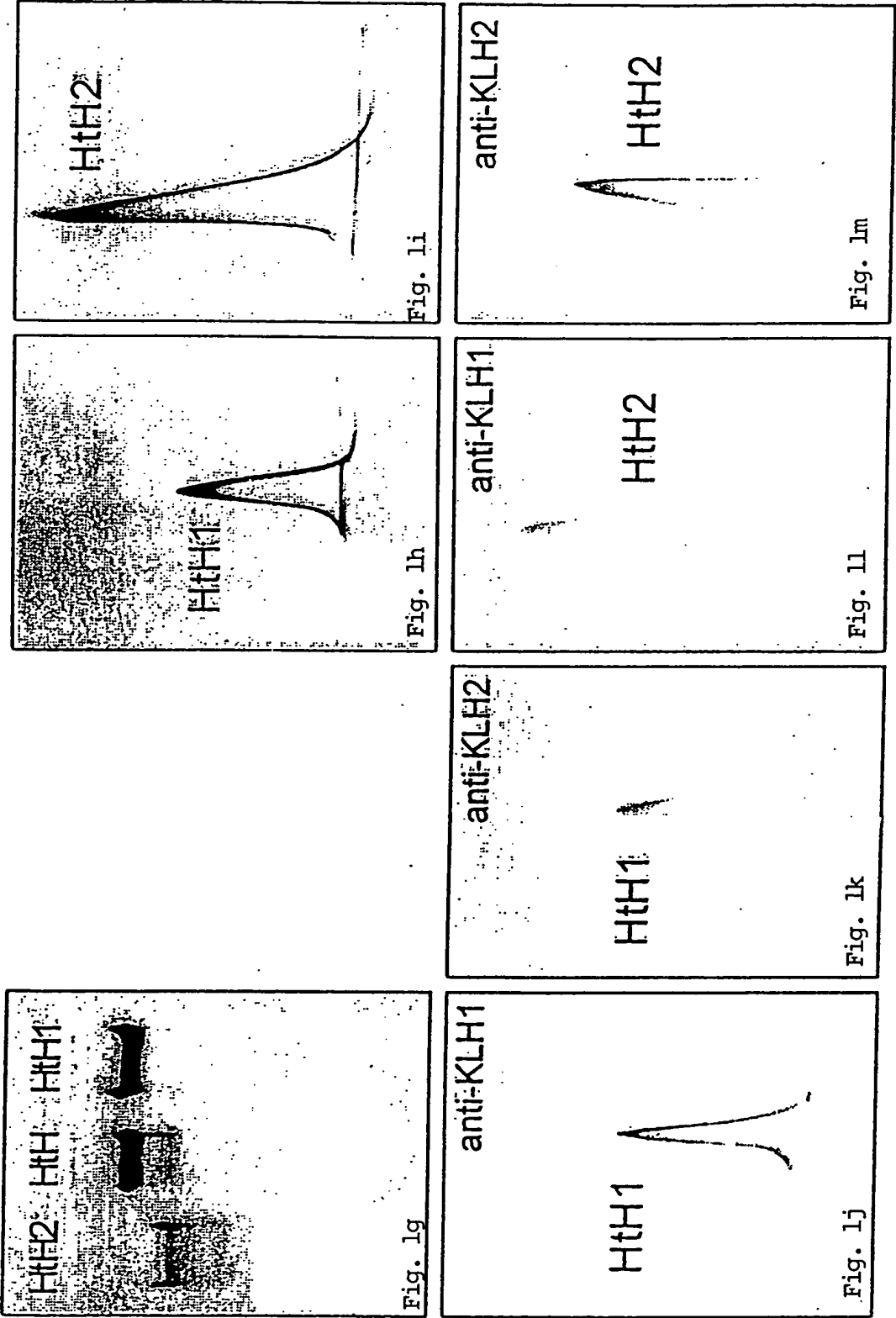
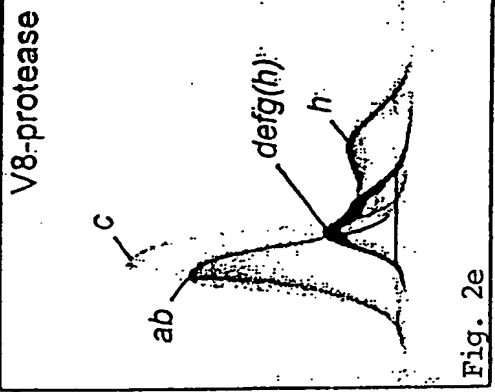
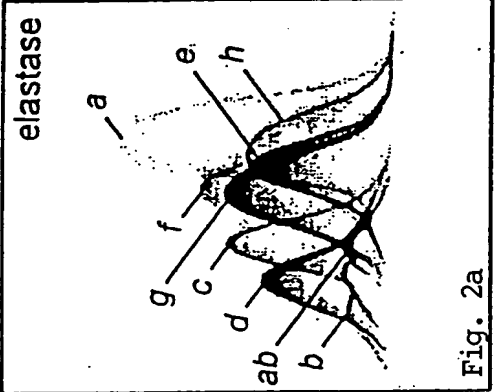
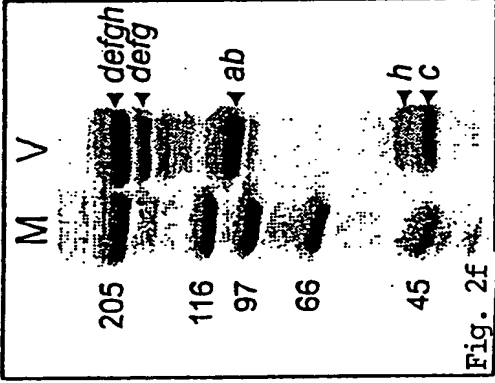
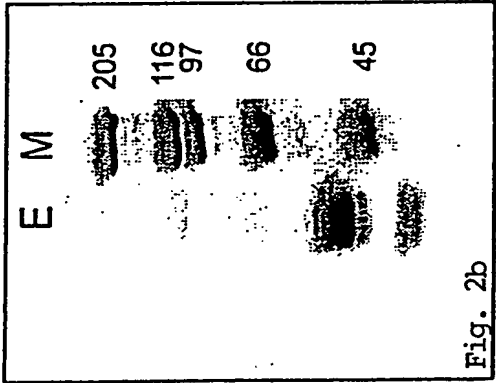
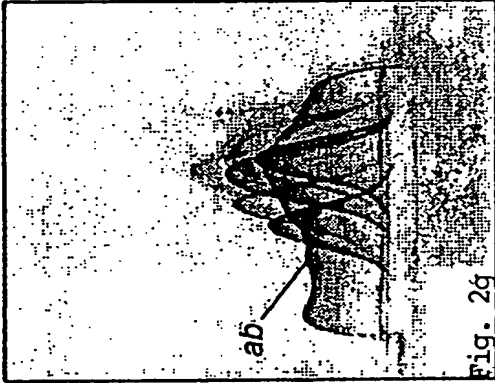
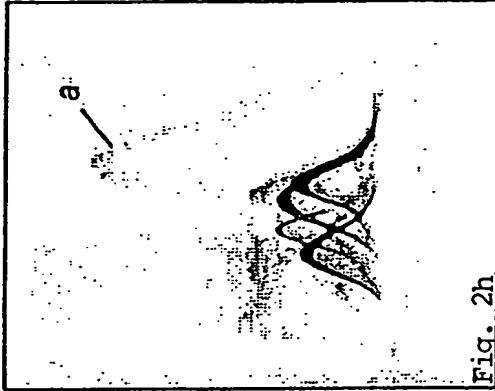
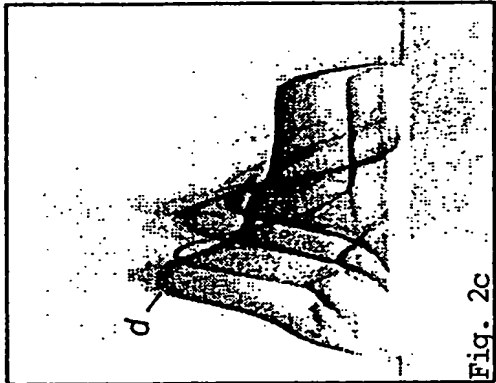
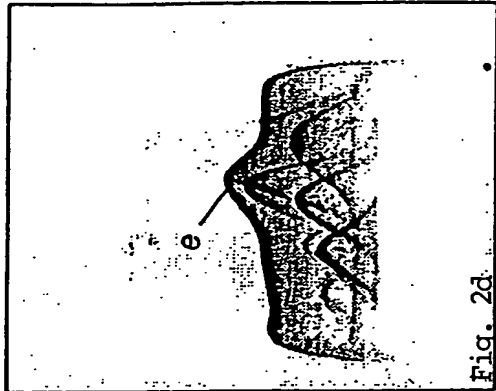


Nucleic Acid Molecule Comprising A Nucleic Acid Sequence Which Codes For A Haemocyanin And Comprising At Least One Intron Sequence (Serial No. 10/049,988; Inventor: Markl Jurgen)



Nucleic Acid Molecule Comprising A Nucleic Acid Sequence Which Codes For A Haemocyanin And Comprising At Least One Intron Sequence (Serial No. 10/049,988; Inventor: Markl Jurgen)



Nucleic Acid Molecule Comprising A Nucleic Acid Sequence Which Codes For A Haemocyanin And Comprising At Least One Intron Sequence (Serial No. 10/049,988; Inventor: Markl Jurgen)

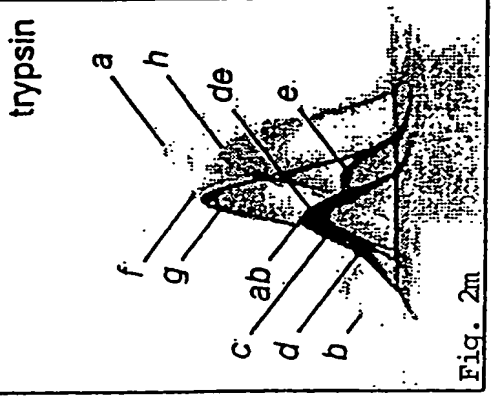
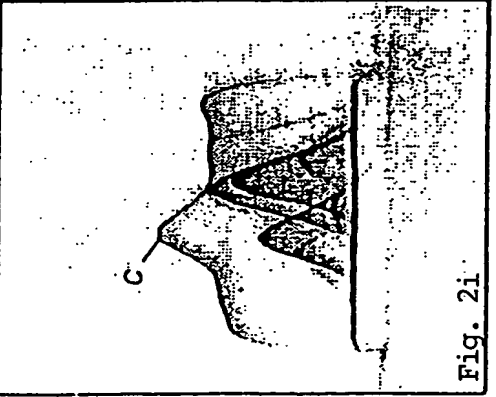
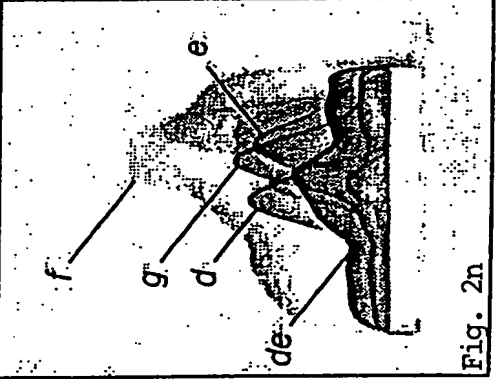
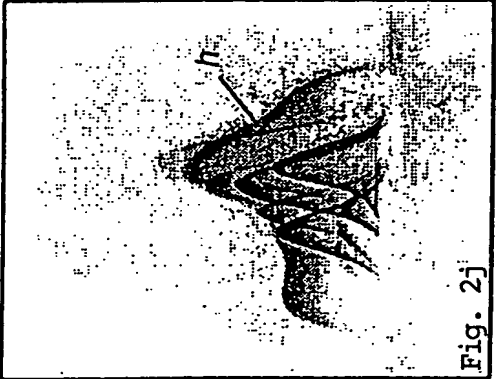
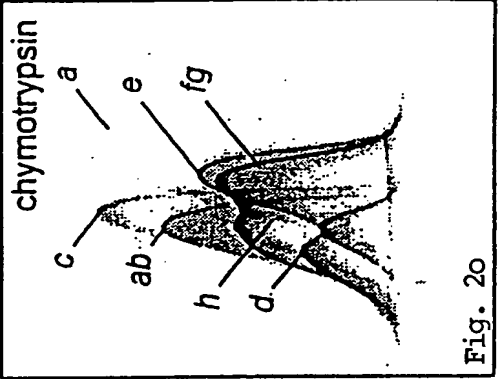
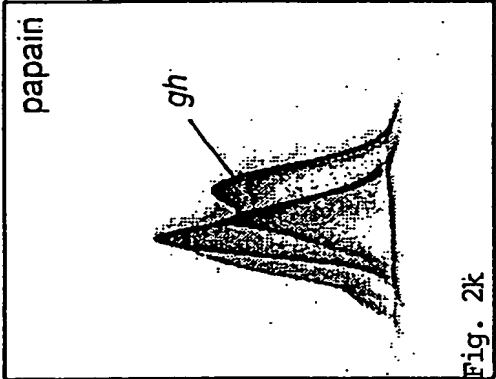
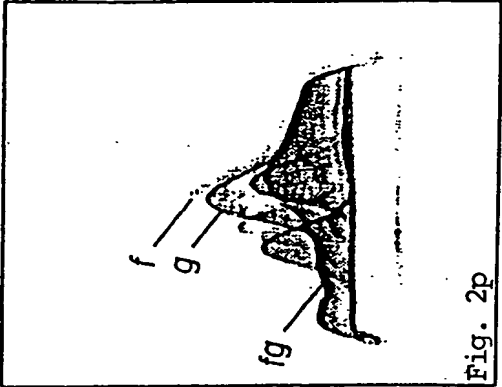
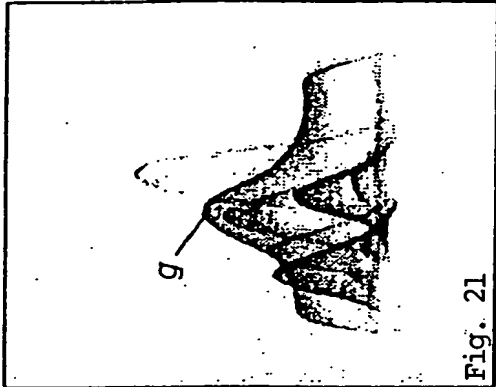


Fig. 3a

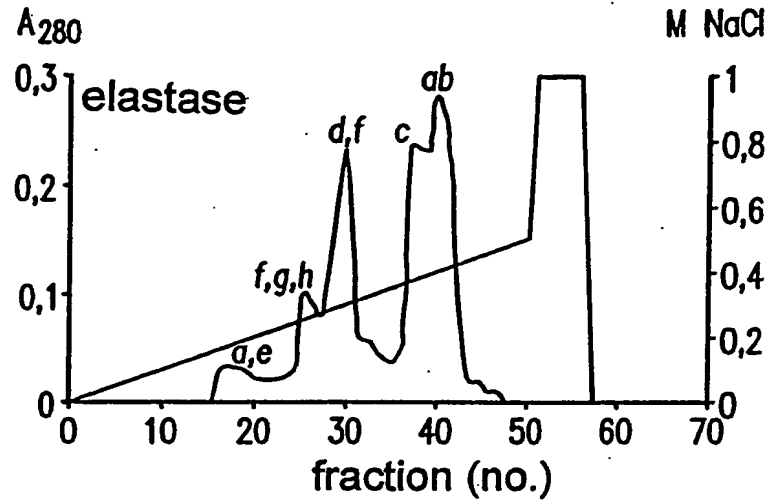


Fig. 3b

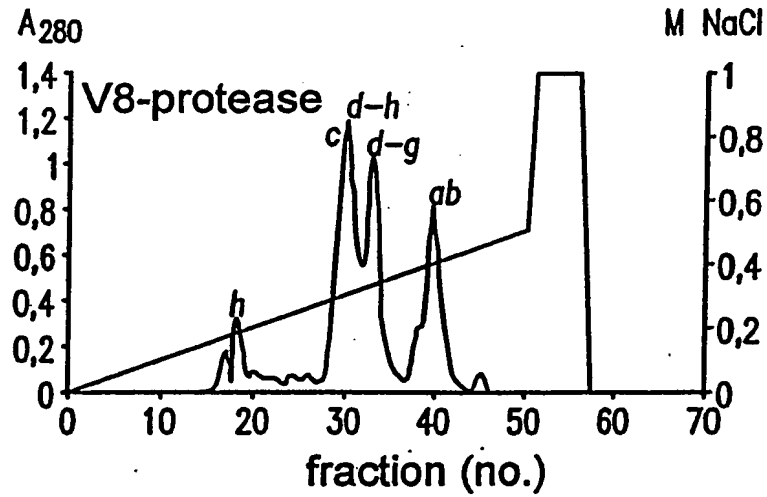


Fig. 3c

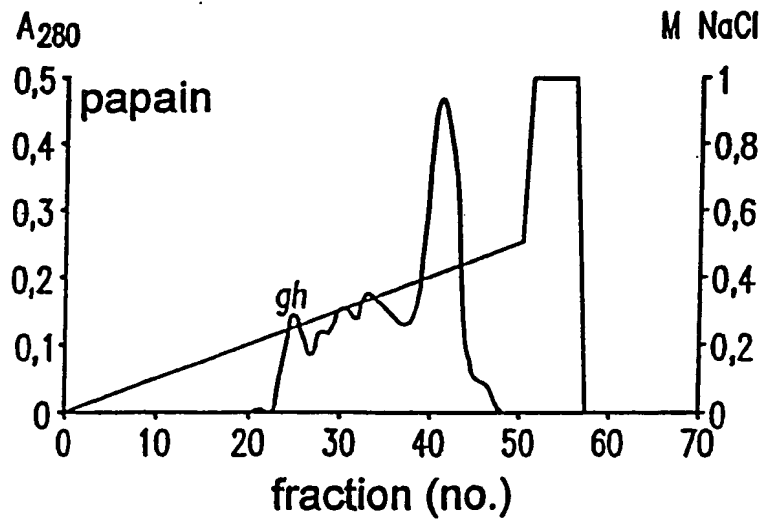


Fig. 3d

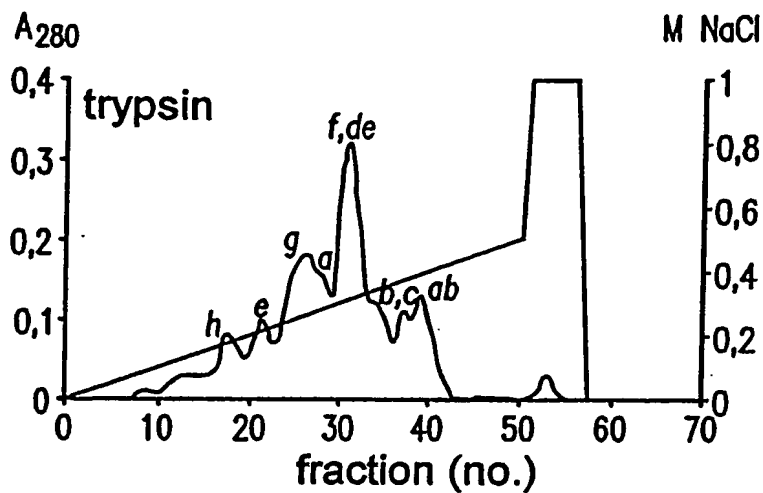


Fig. 3e

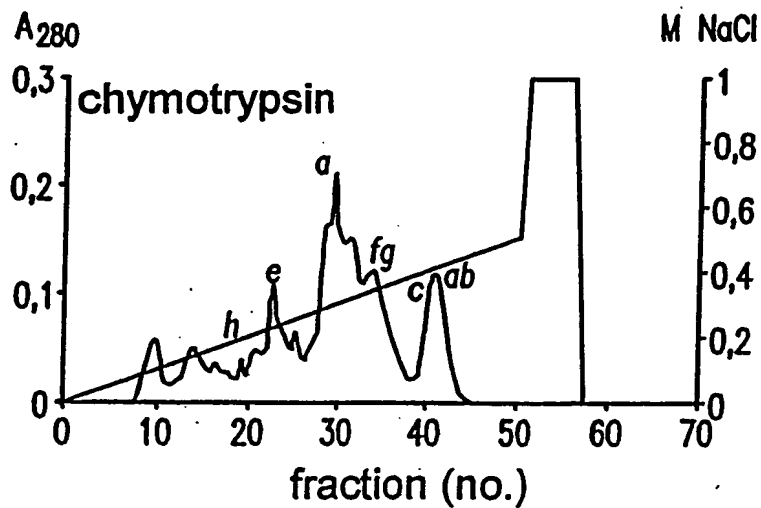


Fig. 4aGenomic sequence of the HtH1 gene

SIGNAL PEPTIDE SEQUENCE 1S-1 (1st part)

GGCTTGTTTCAGTTTCTACTCGTCGCCCTTGTG

INTRON 1S-1/1S-2 (SEQ ID NO:109)

GTAAGTCAACGTCTTTGTTTTAAGTTTGATGCATATCTATCATTGCGTTTTAAAATACCA
TTACAACCAACGTGTCTCTATTGGTCTTCACCTGTTTAACGTATATATTGTTTTTAATGT
GAAAATCTGAGATTATTTTCATTTCCGTCAATATTCGTAAAATACTATACAAATAAAATT
GCTTCAGCCTATTGCATTGGCAGTTTTTCGCAGAATAACGAGGGAAGGCGTACATAAAATA
TAAACCAAGTGTATATTCAAGCATGTTTATAATTTCTTTATAGATTATAACATCATATCAA
AACACCAATCTGGATTTAAACCCGTGAATCCAAAGTATACCAATTAACGGAACCTTTATCA
TGTTTTTATCAAAGGTTTTAGATGAGGGTAAAGAAGTCCGAGCTATATTTTGCGATATCAG
CAAAGCCTTCTATCACGTCTTGCACACAGGGCTGGTATCTAAACTCGAATCCACAGGAAT
AAATATTTTCAGCCGATAGAGAACAGTCGGTGGCTATCATTGGTCACAAAACAAGTCCAAA
ATCTGCATTAGCCGGTGTTCCCCAAGGCTCTGTCTTGGGGCCACTATTATTTCTCACCTA
TATAAACGATTCAACTAATGGAATATAAAGCAACGTAAACCTCACCGCAGATGAAACACT
AAGTTATAGACAATCCGTTTTAAAACCCAGCCACTGCTTAATAATGACTTAGGCCGTCTTT
CAGACTGGGCTAGTAAGCGGCAGGTTAAATTTTACCTTGAAAAGACAGAAACCATGGTAT
ATTTCAAAAACACGAATGCAAGTCCTAAACTTCAACTACTACTTGTATGATACTGGGATTT
CTAAAGTGTGTGAACAAAAACACATTGGCCTGATCCTACAAGATAACCAGACAGAAACCA
TGTTTTTTTTCAATAACACGAATGCAAGTCCTAAACTTCAACTACTACTTGTATGATACTG
GGATTTCTAAAGTGTGTGGTGAACACAAACACCTTGGCCTGATTCTGCAAGATAATGGAA
AATGTCAGAAACATAAGCAAGTTGATGTGGGGTTTTCTGGGGGTGTGACAACACCGAAA
GACCCTGCAACTAATGTTAGCTCAAAGGGTTTTACACCCGGTCACAAGTGGGGATCGACC
CAGGCACCTTTTGCCTTTGACAGCTCGCCTTTCAAAAAATCTCAATTCGAAAACGAAATC
TAATAATTTTCATGAGCGATACAACCGTTTTTTCATAATGCTGTGGTACCGCATACTGTGGA
AACATCTGTCTACCCATTTGGTAGTCCCCCATAAAATGTATTTATGTTTATAAACACAAT
GTTTATAGGGTTACAGTTAGAAGAAGCATTTCTATTGGCTAATGTACATTGCTTGTTTTTT
ACTATTGTGCAAAGGCATATTACAGGTCTTTTAGGAAATTAATACTGTTTAAATCACAT
ACACTACCGGTAATCCTATTATGCTTATCCTGCCAACATTCTGCCCAAGCAAACGCATGA
AAGTTAAAGCTGAGTGTAATAACTGATTGCTGTGTTACTTCACAACCAGTGGACTGAAT
ACAACCATGTTTTTTCTTGAAAGTCACAAACATCCAGTCGGTTTCTAATGTGTAAAGTTT
CTAGTTTCATAAAGAGCATGACGTAATGGTGAATAGGAGTTATCAATGTTTCTATCTAAT
GACTCCTAGTTCGTTACTTTTTTAATAAAACATCCATGTGTTTAATGTTTGGCCACAGAT
ATAACAAGAAAGAAATCGGATAAAATCTACATTTTGACCAATCGGAAGGCTGCCCCCTCC
CTAATCCTAATCATTTTTGTGCCTCAAAACATACTCAACCAGACATTTGAACTATGTATA
TATCAGAATGAAATGGTAACAATAAATTTGTATGTTGACCAGACAGAATTAGGGTGAATC
TGAATACCAACTATTGTACATATGAATATGGATAAGCTCTGCGCGTGCGTGCGGGCGGT
GTAGTGCGGTGTGTGTCTGT
TGT
TGCACAGACATGTGGTTGAGACACACTTGATTTCAGTGCAGGATTATGTCCTTCAACCGAG
TGTAAGTCTTTAAGTGTGCCTGGAAACAAAAAATGCGTTGGGTTGCATCGCCTCTGTAGC
AAGCTTGGACGCGTCACGCAGCTCTGATACCACGTATTGGCACCATTGTTTCATCGGTCTC
ACGCGAATATTATGCTATGTGTGGCGTATCATAACCATAGGTTGGGAACGTTTCAATACTG
TACCGAGCTTGGGCGTGTCACAAAGCTATGATAAGATGACAACACGTCTTGGCATCTTGT
TTCCTCGGTATCACGCGCTGTTATGCTATGTGTGGCTATCACACCTTAGGTTGGGAAAGT

Fig. 4b

TTCCACATTTTCCAGCCTCGTACATGTTTCCTTTTGTTCCTTAGTTATCAGCATAC
CGTATATTCTATATTTAATGAGCATTGTATTTTCTACAG

SIGNAL PEPTIDE SEQUENCE 1S-2 (2nd part)

GTGGGGGCTGGAGCAG

INTRON 1S-2/1A-1 (SEQ ID NO:110)

GTGAGTTTCTTAACATTGTCATGGTACATGGATATACGCTCAGTGGGAAAGCAGGATATC
CCCTTGGTTCAAGTATTCACTTGTACGCCAAGTGTTTCGATTCCCAACATGGAATACTGT
CATATAGTAAATTGATACACTACTTACATTTAATTCTCCACTAAACGTCAACGTCCTTTA
CTTCATGGCCACATGGTCCGTATTAGTGAGTGAGTGAGTCAGGGCATAAGTATTTAACG
TCAAATCAGCAATATTTTCAGCCATATTGTGACAAGAATTGAATATAAAATAATTATACTTA
TAATGCTTATAAATATAAATTATATAAATACCTATAACTATAAATTAGTTATACTAGTAT
TTATCAAAACATATTTGCCACGACACTGCACGCCGATACTTCAAGTGTCTTCACCTCAAG
CGTGTAACCTCCTCATACTCTGTAATAAGTATGTACACTAAGTGAGTGCTATCATCTCCAT
GCTTCATTAGTTTCGTCAGATGCGTGTATCCATACGAGTACATTCAGATTATGGGATCCA
GAGCTTTCTTATCTCAAGTATTTCCGATTGTAAAGCCATACTACTTCCCCAATGACTGAC
GAGACAGATGGCAACCGTTCTTTCCCTCCTGACTAGGTGAGTGCCACTGATAAATCATTAT
GCCTTTAACATTAGGAATGTTAGCAGTGACATGTTTCAGAATTGCGACCTTATGGTTGT
AAAGATTACAACTTTTACAACCTTACTTGAGACAGGTTCCATATGTTCGTATCTGAAATAGT
GTGAAGGTATCTGATTTCGATGCAATACACAGACATATAAACATATTGTTCGCCCTGCTATT
CCGGAAAGGTCATTTTGTATGTAACGTTCCCTAATGGACACAAACGGAATTATTAGTTAA
ACATACTCAACAAAACCTATGTTATTTTGCAATGGGTAGCACCGAAATCTACCGACAGTGG
TTCGTAAAGTAGAACATTCTGACATAAAGAAAAATCATTGGCTTTAAATATATGCAAGT
TACTTGTCTCTAACAACCAGTTTTATACACATTTTCAGAGAACGGGGAATCCGCGATGACA
ATATCAACGAGTATATACAGAATATATAATTAAAAACGATGAGTGCCTGGCAAGGGAAAG
AGCGAGATTTGCCAAACAGGGGGGTGGTGGTTGAGCTTGAATCGTGGAGAAACGTAGATTG
AAAGACAAGATGACATCTAATGATCCGAAAATCAAACACAGGATTAACCTGGGATGCAGAA
GAATGAATATCTCAAGCATAACATGCAACACTTCATGAATGCATCTCAAACATTTTTCGTCA
GATCGGATGCATGAAGATTTGTAAAGCAATGGTTTAAATTGTCCCTAAACGTTTAGTTGG
AGATGTATGAGGCTAGGCTGTATGTTGAACGAAACCATTTAACATTGTTGTTTCATGATTA
TTTAATATTTTTTCATTTTATAGATGTACAATAAAATTGGAACTAAACATTTCCCTTTA
TTGTTTTGTATTTACCTGTTTCATGGGTATGTTTTGAAAGATCGTGATATTTAGTTGGCAT
TCACAAGTTGGAAAAAGGTCACTCAGTTTTCAGTTTCAAGTTTATGTAACCTCTTTATCTGA
CGCTCCAAAATATGTATAGCCTTGTTTCATCTGTTCGGTATGTGGATATTCCTACTTCAGGG
TAGGGTAGCATTAATACTTACAAAACATAACGTGTACCAGATTTTCAGTCACCTCAGAGAT
GATAATGCATGTTCGATATGATAGGTCAAACTTTTCGATATCAATCACAATGAACCTATGG
ACCTGAAATCGGAATGATACGTTACACTTTAGAAACAATTCACAAATATGACTGTCACCC
TTTCAGGTAATAATGTTTGACGGACTACGATAGTGCTGAACAGCAGGAGAGGCAACATGG
TTCGATTGTGAGACAGGTTTAGTGTATTTGTTTGCGAATTTAAGGTTCTGAATCACAATA
GACACGGTTCAGTTAATGGATAAACCAATCATTAGATAGATAGAGATTAGTCGCGATATT
GCTGGGATAAAGCTTAGTGGGACGTTAAGTCCCATCTCAATCTCTCATTTTTTCCAAA
ACAGTTTTTAATTCAGGCTCATGACAAGGTCGTAAGTTGCAAAGGATTCTACTTCAAGCA
GAGATGTCTCATGAATACAGTACAGGGTTTTTGAAGTTTATCCAGTGCAGCGCTGGCACC
ATCTCTGCATGCGAATTATACCATCCATGCCGCTCTAGGCTATTTGTATTAAGTCTGTAG
AATTAAATTCGCGAGTTGCAAATACTGCTCACCATTATCTGCCTCAACCCAGTTTGGGTA
CATCCAAAGTTTTGGGAGATTTTATTCGAGAAATCAACCTGAGATGTTGAATCGGGAGCT
GCGCTTATTCAATGGTGGACTCGGAAGGGAAGTAACCGCTGATGAGGCAAAACAATAACG

Fig. 4c

CAAACATATGGAAGTGGAACCTCTTTGAACCAGTATTATGTTTGTGTGGACATGTATGTGT
TAATTTGACCATTTCGAACAACCTTTACTATTCTATTCATAATGTGTTTAGATTTACATTTG
AATTAAAAGAGATGAGTTTAAAGATATTAATATTTTCCTTTTATAGTCTGTCTGATTGTA
GGGCAATATTTATGTATGTTTCGTTTCATTTTTCATTTATCATTGGAAGGTATATCATAA
GATTATTATTATCATTCTTGAAGTAATGTATACATATATATATGTCTTGAGTAGCTTATT
TTCAATTTATTATCATCCGTCATCCAATTTTATTTTACGAAAGTATAAGAAATAACGAGA
GAGAGAGAGAGAGAGAGAGAGAAAAGACAGAAATGAAGTTAGGAGATATNAGTTATCAAGAA
AACAACAGTTTGAATTTTTTGTGTTAGACAAGATATCATATCAATAACCTCGCACTATTAC
GGGAATAGGCGGGCGTTCCATATGCACAATGAATCGTCAGTTAAAATCAACATTAAACTT
AAAATACTCCTCATATTTAAAGTTGATCTACCTCTTGTATTATTGTAGACTATTAGACAG
AAGTCGACAGTGACACCAGCAACCAGATATCATACCCAGACTTAAAAAGCTGTTTCCTTG
ATGTTTCAATTTATTTCCATTTCCATTATTTCCCTTTTATTGGTTTCCATTTATCAAACCTT
ACCATCTGCACCAGTGAGGAGATTGATATGTTGTATTTATTTATATTTCTTGTACTACAAT
ATCAAGAATGTATAGGAGCTATTCCCTTGTTCCTAAAACCGGATAGATCCATAATTTCCAT
TTTGGGATAAATGGAACTAAACACAACCTTTTACAGTAAACACGAGTGAGCAAGTTGAGT
TTTACGCCGTTTTTTAGTAGTATTCCAGCAATATCGCGGCGGGGGACACCAGAAATGGGCT
TCACACAGTGAATGCATGTGGGGATTTCGAACCCGGGTCTTCGGCGTGACGAGTGAACGCT
TTAGCCACTAGGCTACCCACCGCCTATTTATAGTTAAGACGAATACTTTTCTCAAGCCT
CAAATATGTCCATTCTAGAGAGACTGAATCTGATCCTGAATCTGCGGACCGGTCTTGAAT
ATCATCCCACTAACTCATTGTACAAAGTACCTGTAGATTGTGAGTTCAAAGACAGATTTT
ACAACCCCTATTATATTTTGTCTGCTCATTAAGATATTCAGACTCACTCAAACCTGCTAAA
TGATTTTAACTCTACTTTGAGATGTTTTAACTTTTATTTCGATGCATTTTTTGCGTTCTGCG
TCCTGTATAAAGGTAAAGCAGGTAAACTAACCTAACCTGTTGATTTATTTTCATAGTTTTG
CGATCAGATTGAAACCGGAATGCACAGTGAAGTGTGGCATAACATCTTCCACAGAGATAC
TGGATACTAGGTGGTACAACCGCATTGGCTTTGTGAAAGGATATTAGTGTTTTATGAGAC
TGACTCATGTTTCAATGCTTAGAGCGGAATGATCTCGGTCTTCATGAAAAATATTGTGTT
GAAGTAACCCCCCAGTCCCTAACAGAACGTGGGGAAAGCAGATGGATATGCCAAGACATC
TTCGCATGGTGTGAAGATGATCGTTACAACATCTGCAGAAAAAGTTATTTCTGTGAAGAA
TATGCCAAAGCATCACTGTGAGTGTGTTTGAAGATGTGATATGGCAACACGCAGCGTGTA
TTATGCTTTGTGTGTATTTCTGAAGATCCGTATGAGCATGGCGCCAACTATCAGTTAAA
TGGCTATGCGAAGATCTTCCCGAGATGGTAAACACATATTTTGGCCATTTTCTTTGTAAG
TGGGCGACACAGAAGATCCCCCTGATTGTGTGGATGAGGACACAAAAACGGGTCCCCCTT
CCTTTGCTGATGCTAATGACGCCCTGGAAACATTGAAAGACTTCTTCTCCAGCAAGCAAG
CCACCAACCACAAGTTGTATAAATCGCTTGCGGACTTGAATACGGCAGTTGGACAGATAC
ATACAGCCAGAGAGGGCCGAACATAAACATCTAAACATGGAAAACTGTAAAGACAGGCT
TTGTTGTACGACGTACGTAAATTCATTGAATGTTTGAAGGTTAGAAAATTATTAATCT
TTGAAACCTCGCTCTGTTTGTGTTTGTATTGTCCCCACATTTGCAAATGGTATCCAAAA
GGGCAGACACATTTGTTTTAATCTTAGCCAGGTTCAATTTAGCCTTGCGCCAGACTCAT
TGTATCTGGTGAAGGCTATAGGTGGCCACGTCTTCTAAGATGCTATGCTATTCTTACCAG
AATCCAATGTAAAGAGTTCAAACGCATGGTTCGCTTTGATTGTGATTCTTTCTTAGCACC
TCTCTCTACCCAGAGTTACCTGCACTGCTCCTGACTCACAATAAGCTGACGTGCTGTC
ATATATGTGCAACATTGTATACGTTGGCGTTAAGCCCAACTCACTTCCGCTGTCTTTTGG
CAG

DOMAIN 1A-1 (1st part of domain a)

ACAACGTCGTCAGAAAGGACGTGAGTCACCTCACAGTTGACGAGGTGCAAGCTCTTCACG
GCGCCCTCCATGACGTCACTGCATCTACAGGGCCTCTGAGTTTCGAAGACATAACATCTT
ACCATGCCGCACCAGCGTCGTGTGACTACAAGGGACGGAAGATCGCCTGCTGTGTCCACG
GTATGCCCAGTTTCCCTTCTGGCACAGGGCATATGTCGTCCAAGCCGAGCGGGCACTGT

Fig. 4d

TGTCCAAACGGAAGACTGTCGGAATGCCTTACTGGGACTGGACGCAAACGCTGACTCACT
TACCATCTCTTGTGACTGAACCCATCTACATTGACAGTAAAGGTGGAAAG

INTRON 1A-1/1A-2 (SEQ ID NO:111)

GTA ACTACAAACGTCGTC CCCATTTCATACAGGAGAAATATACAATTGTGTTGTAAGAGCGG
TATACTGTTTGCCA ACTGTGTAATTGAAACGTTGATGATGGTGTCTTTGTATTTCAATTT
GTATGCACTTAGACATGATCAATGTTTCTGATGTGTCAAGGATGTTTCGGTGTGTCACTTT
CAAAAGATCAAATTCATATGACGTACACAGAGCAAGAACCAACAGTAAGAAGTCTGTATG
ACTTCGCTCTTAAAAGCAATGGAAAAATATTTTCACTTAACACCTAGCCCATATCACGC
ATATTAGATTATTCAAGCGATGTCAACATGTTTTTAATATCAATCTCATGGTTCTGATAT
TACCGGAGACATGCAACAGGCTGCCATTATAGCCAGGAAATCTTATGAATATGTGCATAT
TTTTTCTTTGATTCTGTATGACGAGAAATATTCGGAGGCAAAGATTGTGTTTTCAGAACA
GAATCAGGGTATCAGTGACATCGTCACTGCATGGCTACAATATTGCTGATGTGACTGTTT
CTCCAAGGATTTTCATCTCACTGTCTGTACTTTGAATCTACAAATTCGTATTAAAGTTAT
GACAATTTTACCCCTGCCTATTTGTAAACGAAATATAACATGAGTGTTTATGCTGACAG

DOMAIN 1A-2 (2nd part of domain a)

GCTCAAACCAACTACTGGTACCGCGGCGAGATAGCGTTCATCAATAAGAAGACTGCGCGA
GCTGTAGATGATCGCCTATTTCGAGAAGGTGGAGCCTGGTCACTACACACATCTTATGGAG
ACTGTCCTCGACGCTCTCGAACAGGACGAATTCGTAAATTTGAAATCCAGTTCGAGTTG
GCTCATAATGCTATCCATTACTTGGTTGGCGGTAAATTTGA

INTRON 1A-2/1A-3 (SEQ ID NO:112)

GTAAGTTTGGTTTACAGTTTTCATTATAAAAACATAGCAGTTTTAAGTTTAGGGGCAGATT
CTAATCTCTAATATTCCTTTCAACTCACTTTATTGGTGCCTTCTTGGAGTGACATTTAGA
AACTAAGACAAGAGGAAGATGAACAATGTTTGTAGGGATAGACAGCTTGGATGCAATTTTC
GGACCAGATTCTAACAGCGTCATGAAGCAAGTGATACACAACGTTATCAATAACGAGAAT
ATACACATAGATGGTTTGAAGTTTATAAATGAACTATTAACGGCATTGTGGTTATAGACAG
TGAGGAAGACGCCAGATAGACAAAGGGTAGGGGCCCTTGGTTAGATAATGAGAAGTTGAAG
AGGTGTAATAACTTAAATCTCTCTTGACTATTGATTGTGTCTAAGAGTTTTCTTATCTTA
CAGTCGGCCAGTTGGGTCAAAGATGGTGTGATTCCGGATGTGCTTTGTGTGTTCTGCGATG
GCTGATTTAGAGTCAGTTTACTTCAGATGAATGAAGTTCCCCGATTCTTATGTTTAAGTT
TGTTTACCTACGCATGAAGACATCACCAGCAGGGTCGTCTTTATTTCTAGTAGCTTATT
TACAGCAAGCTTGTAACGTATGCTGAATTGCTGTGCCTCTGTAGAACACAGCATCTATGT
TTGCTTGCTTCTTTAGTAGACTGCGGATGTGATGGTTGGTTACCTGGTATGCTGACGAAA
GAATTGTTGACGTGGTGGTTTGCCTTGATGGGTTTCGTTGACTTGGTTTGTGATACTGA
TTAAGGTGACTCTGCTGGGAGGCTTGGAATCTGGGGCCGGTGTCTTTGCTCTCCTGTCT
AGGGTGGCGATTATTTCCCAACCCACTTGTTCCATTACACTCAAACCTGCTATCAATTT
ACAG

DOMAIN 1A-3 (3rd part of domain a)

ATATTCAATGTCAAACCTTGAATACACCTCCTACGACCCCATCTTCTTCCTCCACCACTC
CAACGTTGACCGCCTCTTCGCCATCTGGCAGCGTCTTCAGGAACTGCGAGGAAAGAATCC
CAATGCAATGGACTGTGCACATGAACTCGCTCACCAGCAACTCCAACCCTTCAACAGGGA
CAGCAATCCAGTCCAGCTCACAAAGGACCACTCGACACCTGCTGACCTCTTTGATTACAA
ACAACTTGATACAG

Fig. 4e

INTRON 1A-3/1A-4 (SEQ ID NO:113)

GTGAGACATTATTACACTTCTATTTAGTAGTGGGGGCGGGATAGCTCAGGTGGTAGAGCG
TCGGCCTTCAGCTTCTAGTCTCGCCACAAGAGCGCGCTGGCTAAAGGGCCGGAGTTAGA
TTCCCGCGGGGCGGCAGGCAATATCTCCGAAGGGGAGAACAGTTCTCCAGTCGGTGAAATT
GGGGTGCAATGTTGTACCACTGAAATGCGTGCAGCACCAACCATCCAAATACCAGCCTTG
CCGCGCTGGTCTGACTACATAGTACCACCCGGATTCAACCGGGCTATATAGGTTCTCCTC
CAGCAGTAAATCTGACAGTCGCCATATAGCTGGGATATTGCTGAGTGCGACGTTAAGCCC
CAACTCACTCACTTTATATTTAGTATTCTATTTAGTATCGACGCATGACCATGTGTGGTG
GTCTACTCATCTCAACACGACCGATTAAACGTTAAGAGCTGCCAACATGATTCTCTTTCTC
TCTTTAGCCTCTTTATGCCAAAAGCTATATATTAATGTAGGACCCTACATATATTATTTTC
CAG

DOMAIN 1A-4 (4th part of domain a)

CTACGACAGCTTAAACCTGAATGGAATGACGCCAGAACAGCTGAAAACAGAACTAGACGA
ACGCCACTCCAAAGAACGTGCGTTTGCAAGCTTCCGACTCAGTGGCTTTGGGGGTTCTGC
CAACGTTGTTGTCTATGCATGTGTCCCTGATGATGATCCACGCAGTGATGACTACTGCGA
GAAAGCAGGCGACTTCTTCATTCTTGGGGGTCAAAGCGAAATGCCGTGGAGATTCTACAG
ACCCTTCTTCTATGATGTAAGTGAAGCGGTACATCACCTTGGAGTCCCGCTAAGTGGCCA
CTACTATGTGAAAACAGAACTCTTCAGCGTGAATGGCACAGCACTTTCACCTGATCTTCT
TCCTCAACCAACTGTTGCCTACCGACCTGGGAAAGGTCACCTTGACC

INTRON 1A-4/1B (SEQ ID NO:114)

GTAAGTTGATTGTCTTAATATTGTTTTAATTTTTGCAGAAATTTGATTTTAAATTGTGTA
ATAACAGTACACATTTTTACGCAACAGCAGTCATTATTGTGTGTGAAGATGTCAAACCAG
AAAGGTTTCAATCGTGAAAACAAAAACAATTCTCTATCTGTATACCCCTCAATACCAGTA
TGATCACAAATCTAGGAAATATTACAATACTGCTTCATAGAGTAACTGCTGTTTGTGGCA
GAGCTGGATACGAAGTTTCTGATAGTTCACAGCTACATGATAGTAAATGAACCTGTACAC
ATCAACGGTTGATCATGAAAATTTTGTATGTGTGAAAGTGCTACCTGTATTAGTGAACGT
GCTACCTGTATAACTGAAAGTGCTACCTGTATGACTGAAAGTGCTACCTGTATGCTGAAA
GTGCTACCTGTATTAGTGAACGTGCTACCTGTATAACTGAAAGTGCTACCTGTATGACTG
AAAGTGCTACCTGTATTAGTGAAGTGCTACCTGTATGAGTGAACGTGCTACCTGTATAA
CTGAAAGTGCTACCTGTATGACTGAAAGTGCTACCTGTATTAGTGAAGTGCTGCCTGTA
TTAGTGAAGTGCTACCTGTATGACTGAGCGTGTTACCTGTATGACTGAACGTGCTACCT
GTATTAGTGAAGTGTAATCTGTATGAGTGAAGTGCTACCTGTATTAGTGAAGTGCTA
CTTGATTAGTGAAGTGCTACATGTATGACTGAAAGTGCTACATGTATGAATGAGAGTG
CTACCTGTGTGACTGAAAGTGCTACCTGTATTAGTGAAGTGCTACCTGTATGACTGAAC
GTGCTACCTGTATTAGTGATAGTGTCACTGGTACCAACTGGATGTTCTCACTTCTTTGGC
GAATATCTGGGCTCAAACAGTTTTTTCAGTATCATAGTCGTATCAGTTTGATTTGTATGT
GCAGTGGAATCATTTTTCGTCAAATAATCAAACCTGGTGTGAACTGGCGTTACGTTTTTA
TGGTTGTAAAACAAATTCTGTAAGTAAAGATATTTTAGGGATATCTGTATGACATGAACT
GAATTGCTTAAGGTTAGCATGCCATGACAAATTGCTGAATGTCTGAGGATTGGTGGAGCA
ATAAATCATTATTAAGACAAAATCAGAAACGTCCATTTTCACTTTTAAACAGTGTATCTG
TCTGAATGCCCCCTACTTTTTGGAAGAGTATATATGAATTATCGGCAATATAAACGTTA
AATGGCAAATGTGCGGCATATGTCAGGACATTATTACCGCAGTTTATAGTCATATTTACC
GGGTCTAGGACAATTGTCACCCCGACAATTGCCACCCGGACAATTGCCACCCAAAAATAA
AATATACGTAAACAGAAAACAAATATTGCTTTCAGCCTTTATTGAGTTAGATAATGACAT
TTATGTTGATAAATATGTCGTTTGATAATAATAACAATAATATAATATTACAATACT
GCAATAGTACTATCAGTACTTATCATTTTTATCACAGATTATATATAGATTCTAGAGTCCG

Fig. 4f

ATGTTGTAGGCAACACTTCGTCGGTAGGCCGTTAGGTAGTTATCATTAGGGCTGAGTATT
GCGCCAAATTTTCGTATTGCTATATACTGCGATACACGGTTACCTGTTTTGCAATACGTAA
ACTTAGGCAAATATGACAGTTTTTCCATGATTATTTTACGTTTCAATGCTTAAATGGT
CTTATCTGTTATCTCCTTGAAGGTTTAATAAAATAACAATAAACATAAATCATTATTGAA
AATTAATGAACAAAAGTAAAGCGCTTCTCAGTTACCTTAACCTAACTTATTTATGAATGG
GATTACTATCCAAGAATGTGAAATTACAAACACCTTGGGATAACACTGCAAAACGACTG
TTCATGGGACGGACATGAAAAAGGTGAGTCCCATGTTAAACTGTTGAGAAAGTTTCCTAT
ACTGTTTGTCCCGAAAAAGGCTAAAGACCATGTACTAATCAATTATTCTATCTATTTTCG
ATTACTGTTCTCATATTTGGGACAACTGTGCAGATCGGTAGCATCCAAGCTCGTCTAAAT
CGGTTTGATAAACCTTGTCAAATAACATGTTGTCTCAACATCCAAGCTCACCTAAACCTT
GTCAATACCTGCATCTGAACAAATGTATATTTAAGACGATAGCATCCAAGCTCATCTTTA
AAATGAATATTTTCTCTTTTTCTACCAAAACATTATTTGGTTGACAGTTGTCCTCCCTAT
TATAGTAAAAAGAACTGGGTGGCAATTGTCCTAGGTGGCAATTGTCCGGATGGCAATTGT
CCGGGTGGCAATTGTCCGGGTGGCAGTTGTCCAGGTGGCTATTGTCCTGTTCCCATATTT
ACGTATCCCATTTTCTGCTCTGTAATTTTAAATAAACTCACCTGCCTAAGGTAAGACGAC
ATGTGTCACGTGAACATCGTTTGGGGGCAAGGGCGGAATCCCTTCGTTGAAAGTAAATGA
ATACTGTACATAGAGATGCGTATCTTGAACCTTTATTAGCTTTGATATTGTGCTTAATA
TTACATGAATGTATTTCAATATGTAATTATGTGTTCAAATGAATGGTTGACTTGAATGGT
TTTATTGCTTTATATGCTACATCAACATGTGTGTTTCTTTTCATTTGAG

DOMAIN 1B

CACCTGTGCATCATCGCCACGATGACGATCTTATTGTTGAAAAAATATAGATCATTGGA
CTCGTGAAGAGGAATACGAGCTAAGGATGGCTCTGGAGAGATTCCAGGCCGACACATCCG
TTGATGGGTACCAGGCTACAGTAGAGTACCATGGCCTTCCTGCTCGTTGTCCACGACCAG
ATGCAAAAGTCAGGTTTCGCCTGTTGTATGCATGGCATGGCATCCTTCCCTCACTGGCACC
GGCTGTTTCGTTACCCAGGTGGAAGATGCTCTTGTACGGCGTGGATCGCCTATCGGTGTTT
CTTATTGGGACTGGACAAAACCTATGACTCACCTTCCAGACTTGGCATCAAATGAGACGT
ACGTAGACCCGTATGGACATACACATCATAATCCATTCTTCAATGCAAATATATCTTTTG
AGGAGGGACACCATCACACGAGCAGGATGATAGATTGAACTGTTTGCCCCAGTCGCTT
TTGGGGAGCATTCCCATCTGTTTGTATGGAATCCTGTACGCATTGAGCAGGAAGATTTCT
GCGACTTTGAGATTGAGTTTGTAGTTAGTCCATAATTCTATTTCATGCGTGGATAGGCGGT
CCGAAGATTACTCCATGGCCACCCTGCATTACACAGCCTTTGACCCCATTTTCTACCTTC
ATCATTCCAATGTCGATCGTCTATGGGCAATCTGGCAAGCTCTTCAAATCAGGAGACACA
AGCCATATCAAGCCCACTGTGCACAGTCTGTGGAACAGTTGCCAATGAAGCCATTTGCTT
TCCCATCACCTCTTAACAACAACGAGAAGACACATAGTCATTGAGTCCCGACTGACATTT
ATGACTACGAGGAAGTGCTGCACTACAGCTACGATGATCTAACGTTTGGTGGGATGAACC
TTGAAGAAATAGAAGAAGCTATACATCTCAGACAACAGCATGAACGAGTCTTCGCGGGAT
TTCTCCTTGCTGGAATAGGAACATCTGCACTTGTGACATTTTCATAAATAAACCGGGGA
ACCAACCACTCAAAGCTGGAGATATTGCCATTCTTGGTGGTGCCAAGGAAATGCCTTGGG
CGTTTGACCGCTTGATAAGGTGCAAAATACTGACTCATTGAAGACACTTTCTCTCGATG
TCGATGGAGATTATGAAGTCACTTTTAAATTCATGATATGCACGGAAACGCTCTTGATA
CGGACCTGATTCCACACGCAGCAGTTGTTTCTGAGCCAGCTCACC

INTRON 1B/1C (SEQ ID NO:115)

GTAAGTAAATTTACAAAATTTGGTGTTCTCTAACTATCCTAAGTATTCAATCGTTAGCGT
GTACCTATCTGCATAATGCAATACCCTGACTCCATATAAGTATAGTATATTTACTCTGGT
CGAAAACAAACAAATTGAAAACAAGAGTGGACGTGCTGTTATGATTTCTTTTTTCAATCTT
GGTTCGTTGTGTAATGCCACAGCCAGCAATTCAGATATATAGCGACGGTCTATGAATAC
TCCAGTCTGGACCAGACAATCGTGTGGAATGGTTTAGGCACATTATATCAAATTCATTGT

Fig. 4g

TGAAGATATGAGTTATGAGGTCACAATGTTGTCTTGTTACCCCGTGTCAGTAGTGACGTC
ATTTTCATGACTGAAATCTCTTCAACGCCGTTTAGCAATAATAGGCTCAGTAGTATTCAAC
CAATTACAATCAGTAGAAAATTCTCTATACTATTCTTATGTTGCATCCTGATATCCCTAT
GCAAAAATTAGTCATCTAATATAATCATTTTCGATAAAATACTTTGGGCAAACAAATCAAT
GTAACATCTATTTTCTTTTCAG

DOMAIN 1C

CTACCTTTGAGGATGAAAAGCACAGCTTACGAATCAGAAAAAATGTCGACAGCTTGACTC
CTGAAGAAACAAATGAACTGCGTAAAGCCCTGGAGCTTCTTGAAAATGATCATACTGCAG
GTGGATTCAATCAGCTTGGCGCCTTCCATGGAGAGCCTAAATGGTGCCCTAATCCTGAAG
CGGAGCACAAAGTTGCATGCTGTGTTTCATGGCATGGCTGTTTTCCCTCATTGGCACAGGC
TTCTTGCTCTCCAGGCGGAGAATGCTCTTAGAAAGCATGGGTACAGTGGTGCTCTACCAT
ACTGGGATTGGACTCGCCCCCTTTCCCAACTTCCTGATCTGGTTAGTCATGAGCAGTATA
CAGATCCTTCCGACCATCACGTGAAGCATAACCCGTGGTTCAATGGCCACATCGATACAG
TAAATCAGGATACCACCAGAAGCGTACGGGAGGATCTTTATCAACAACCTGAATTTGGAC
ATTTACGGATATTGCTCAACAAGTCCTCTTAGCATTAGAACAAGATGACTTCTGTTTCGT
TTGAAGTGCAGTATGAGATTTCCCATAAATTTATCCATGCACTTGTAGGAGGAACCGACG
CTTATGGCATGGCATCGCTGAGATATACAGCATACGATCCAATCTTTTTCTTGATCATT
CAAACACCGACAGGATCTGGGCTATTTGGCAATCCCTGCAAAAATACAGAGGCAAACCGT
ACAACACTGCCAACTGCGCCATAGAATCTATGAGAAGGCCCTGCAACCATTGTTGGACTAA
GCAGTGCCATTAACCTGACAGAATCACCAGAGAGCATGCTATCCCGTTTGATGTCTTCA
ACTATAGAGATAACCTTCATTACGTATATGATACCCTGGAATTTAATGGTTTGTTCGATTT
CACAACCTTGATAGAGAGCTGGAAAAAATCAAGAGTCACGAAAGAGTATTTGCTGGATTCT
TGCTGTGCGGGATTAAAAAATCTGCTCTTGTTGAAATTCGAAGTTTGTACTCCACCTGATA
ATTGTCATAAAGCAGGGGAGTTTTATCTACTCGGGGACGAAAACGAGATGGCTTGGGCCT
ATGACCGACTTTTCAAGTATGATATTACTCAGGTTCTGGAAGCAAACCATCTACACTTCT
ATGATCATCTCTTCATTTCGCTACGAAGTCTTTGATCTTAAAGGAGTGAGTTTGGGAACTG
ACCTGTTCCACACTGCAAATGTGGTACATGATTCCGGCACAG

INTRON 1C/1D (SEQ ID NO:116)

GTACGTGGATTTGATTACATAGCAATGCTATATGATTTAGTAATTACAACCTCAAGTCA
TGTAGCCGTTTTAGATTGCATTACATCAAACAGCATTGGATTAAATTGGGGGATTGTCCA
GGCCGCATTATGTTGCATTCCGAAAATAGTTTGTGTCCAGTGTCCACGTTTAAATTTAAA
CCATTTTAATCATATTAGGGATAATTTTAATAGATGTTATAGTGCTTTATTTTCATATTGT
TACAGTGGACAGTCACCAAGGACATATTTTACTCTATAGATACACAAACACCAATTAAAA
CCCTGCTTTGGAAAGTCTAACTTTTTCCCCACAG

DOMAIN 1D

GCACCCGTGATCGTGATAACTACGTTGAAGAAGTTACTGGGGCCAGTCATATCAGGAAGA
ATTTGAACGACCTCAATACCGGAGAAATGGAAAGCCTTAGAGCTGCTTTCCTGCATATTC
AGGACGACGGAACATATGAATCTATTGCCAGTACCATGGCAAACCAGGC AAATGTCAAT
TGAATGATCATAATATTGCGTGTTGTGTCCATGGTATGCCTACCTTCCCCCAGTGGCACA
GACTGTATGTGGTTTCAAGGTGGAGAATGCTCTCCTAAACAGGGGATCTGGTGTGGCTGTTT
CTTACTGGGAGTGGACTGCTCCCATAGACCATCTACCTCATTTCATTGATGATGCAACAT
ACTTCAATTCCCGACAACAGCGGTACGACCCTAACCCTTTCTTCAGGGGAAAGGTTACTT
TTGAAAACGCAGTCACAACAAGGGACCCACAAGCCGGGCTCTTCAACTCAGATTATATGT
ATGAGAATGTTTTACTTGCACTGGAGCAGGAAAATTATTGTGACTTTGAAATTCAGTTTG
AGCTTGTTTCATAACGCACTTCATTCCATGCTGGGAGGTAAAGGGCAGTACTCCATGTCTCT

Fig. 4h

CCCTGGACTATTCTGCGTTTGATCCCGTCTTCTTCCTACATCATGCCAACACGGACAGAC
TGTGGGCAATCTGGCAGGAACTACAAAGATTCCGAGAACTGCCTTATGAAGAAGCGAACT
GTGCAATCAACCTCATGCATCAACCACTGAAGCCGTTTCAGTGATCCACATGAGAATCACG
ACAATGTCACTTTGAAATACTCAAACCACAGGACGGATTTCGACTACCAGAACCACTTCG
GATACAAGTATGACAACCTTGAGTTCCATCACTTATCTATCCCAAGTCTTGATGCTACCC
TGAAGCAAAGGAGAAATCACGACAGAGTGTTTGCGGGCTTCCTTCTTCATAACATAGGAA
CTTCTGCTGACATAACTATCTACATATGTCTGCCTGACGGACGGCGTGGCAATGACTGCA
GTCATGAGGCGGGAACATTCTATATCCTCGGAGGCGAAACAGAGATGCCTTTTATCTTTG
ACCGTTTGTATAAATTTGAAATCACCAAACCACTGCAACAGTTAGGAGTCAAGCTGCATG
GTGGAGTTTTTCGAACTGGAGCTTGAGATCAAGGCATACAACGGTTCCTATCTGGATCCCC
ATACCTTTGATCCAACTATCATCTTTGAACCTGGAACAG

INTRON 1D/1E (SEQ ID NO:117)

GTAATGCCATCTTAATACAGTTCGTTTCGTTAAATTATATATGTTCGTTTACAACACCATA
CCTTGAATTGAGGTAATACATCACTTGATATTGATAATGTAATGGTAATTGTTCTTGTTT
GTAAAACCGTTTCTGGGGTGTTTATTCATCTACCTGGTGGATAGTGAGTAAACACAT
TCGGTTTAATATGGGTATCTAATGGACAGTGAAGTGTGCTGGCTAGGCAGATACCTTGGT
TTCTGTGAATGGAGGTAGTAGAAAGGGGTTTTGATGATTGCAG

DOMAIN 1E

ATACCCATATCTTGGACCACGACCATGAGGAAGAGATACTTGTGAGGAAGAATATAATTG
ATTTGAGCCCAAGGGAGAGGGTTTCTCTAGTCAAAGCTTTGCAAAGAATGAAGAATGATC
GCTCCGCTGATGGGTACCAAGCCATTGCCTCTTTCATGCCCTGCCACCACTCTGTCCCA
ATCCATCTGCAGCTCACCGTTATGCTTGCTGTGTCCATGGCATGGCTACATTTCCCCAGT
GGCACAGACTGTACACTGTTTCAGGTTTCAGGATGCCCTGAGGAGACATGGTTCACTTGTTG
GTATTCCTTACTGGGACTGGACAAAACCAGTCAACGAGTTACCCGAGCTTCTTTCTTCAG
CAACATTTTATCATCCAATCCGGAATATTAATATTTCAAATCCATTCCCTCGGGGCTGACA
TAGAATTTGAAGGACCGGGCGTTTCATACAGAGAGGCACATAAATACTGAGCGCCTGTTTC
ACAGTGGGGATCATGACGGATACCACAACCTGGTTCCTCGAACTGTTCTCTTTGCTTTGG
AACAGGAAGATTACTGCGATTTTGAATAACAATTTGAGATAGCCCATAATGGCATCCACA
CATGGATTGGTGGAAGCGCAGTATATGGCATGGGACACCTTCACTATGCATCATATGATC
CAATTTTCTACATCCACCATTACAGACGGACAGAATATGGGCTATTTGGCAAGAGCTGC
AGAAGTACAGGGGTCTATCTGGTTCGGAAGCAAACCTGTGCCATTGAACATATGAGAACAC
CCTTGAAGCCTTTCAGCTTTGGGGCCACCCTACAATTTGAATAGTCATACGCAAGAATATT
CAAAGCCTGAGGACACGTTTGAATAAAGATTTGGATACAGATATGATAGTCTGGAAT
TGGAGGGGCGATCAATTTCTCGCATTTGATGAACTTATCCAGCAGAGACAGGAGAAAGACA
GAACTTTTGCAGGGTTCCTCCTTAAAGGTTTTGGTACATCCGCATCTGTGTCAATTGCAAG
TTTGCAGAGTTGATCACACCTGTAAAGATGCGGGCTATTTCACTATTCTGGGAGGATCAG
CCGAAATGCCATGGGCATTTCGACAGGCTTTATAAGTATGACATTACTAAACTCTTCACG
ACATGAACCTGAGGCACGAGGACACTTTCTCTATAGACGTAACATACAGTCTTACAATG
GAACAGTACTCTCGGGAGACCTCATTACAGACGCCCTCCATTATATTTGTACCTGGACGCC

INTRON 1E/1F-1 (SEQ ID NO:118)

GTGAGTACCTGTTTGCACCTAAGACTTCTGTAGGCTAAAAGTGTAAGAAATATCAATTAAT
TTCAATTCACCCAACTTGAAAACGGTACCTATATAGGTAACTTTTTGTCTACAGTAA
CTGAACATACCTACACATTTTCATGAAATGATCTCTCAATATTTTCCACCAACAG

Fig. 4i

DOMAIN 1F-1 (1st part of domain f)

ATAAACTCAACTCACGGAAACATACACCTAACAGAGTCCGCCATGAGCTAAGTAGCCTTA
GTTCCCGTGACATAGCAAGCTTGAAGGCAGCTTTGACAAGCCTTCAACATGATAATGGGA
CTGATGGTTATCAAGCTATTGCTGCCTTCCATGGCGTTCCTGCGCAGTGCCACGAGCCAT
CTGGACGTGAG

INTRON 1F-1/1F-2 (SEQ ID NO:119)

GTAAATTTACAGAGCTTTATGAAGTGTGTTTCAGAGTGAAGAGACCAAGATATACTTATAC
CCAAAAGTAGCTAGCAACAGACGATTTCACTTGTTCGGACACTTTGTATTATACGTTGG
ATCCCAAGGTAAACGGAAACGTAACCGAGAATCAGTCCGTAAAGTGAGTGAGTGAGTTTG
GGGCTTAACGTCGCACTCAGCAATACCCAGCTATGTGGCGACTCTCAGATTTACTGCTG
GAGGAGAACCTACATAGCCCGGTTTAACCCGTGTGGTATGTAGTAAGACCAGCGCGGCAT
GGCTGGTATCTGACGGACGAAGGGTGGCGCTGCACGTATTCCAGTGGTACAACACTGCAC
CCCAATTTACCCGACCGGAGAAGTATCTCCCTTCGGAGATATCGCCTGCCTTCCACGG
GATTCGAACTCGGTGACCTTCAAGCCAGCGCGCTTCTAGCGGGGGCGATTAGAGGTTNAA
GGCCGACGGCTCTACCACCTTAACCTATCCCCCGGCCCACTCCTGACGGAAATGTTTATA
ATTGAGCCTTTGTTTTCTTATTAACACTCTTGGCAGATTTTCTATAGATAATGGATTCA
CATGTAGACAGTCTCCCATTTGTTGTAAGTGGTAGTCAAGAGTTAGAATCTGAATACATTC
TCCAAGATGGATCAAGGAAAACAATAATTACTTGATGTTGCAG

DOMAIN 1F-2 (2nd part of domain f)

ATCGCCTGTTGCATCCACGGCATGGCGACGTTTCCTCACTGGCACCGGTTGTACACTCTG
CAGTTGGAGCAAGCGCTGCGCAGACACGGGTCCAGTGTGCTGTTCCATACTGGGACTGG
ACCAAGCCAATCACCGAACTGCCACACATTCTGACAGACGGAGAATATTATGACGTTTGG
CAAAATGCCGTCTTGGCCAATCCGTTTGCAAGAGGTTATGTGAAAATTAAAGATGCATTT
ACGGTGAGAAATGTCCAGGAAAGTCTGTTCAAAATGTCAAGTTTTGGAAAGCACTCGCTT
CTGTTTGACCAGGCTTTGTTGGCTCTTGAACAACTGACTACTGTGACTTCGAAGTTCAG
TTTGAAGTGATGCATAACACGATCCATTATCTCGTAGGAGGGCGTCAAACGTACGCCTTC
TCCTCTCTCGAGTATTCCTCATAACGATCCAATCTTCTTTATTACCACTCGTTTGTGAC
AAAATATGGGCTGTATGGCAAGAACTGCAAAGCAGGAGACATCTACAGTTTAGAACAGCT
GATTGTGCTGTGGGCCTCATGGGTGAGGCAATGAGGCCTTTCAACAAGGATTTCAACCAC
AACTCGTTACACCAAGAAGCACGCAGTCCCTAATACAGTATTTGATTATGAAGATCTTGGC
TATAACTATGACAACCTTGAAATCAGTGGTTTAACTTAAATGAGATCGAGGCGTTAATA
GCAAAACGCAAGTCACATGCTAGAGTCTTTGCTGGGTTCTGTGTTTGGATTAGGAACT
TCGGCTGATATACATCTGGAAATTTGCAAGACATCGGAAAACCTGCCATGATGCTGGTGTG
ATTTTCATCCTTGGAGGTTCTGCAGAGATGCATTGGGCATACAACCGCCTCTACAAGTAT
GACATTACAGAAGCATTGCAGGAATTTGACATCAACCCTGAAGATGTTTTCCATGCTGAT
GAACCATTTTTCTGAGGCTGTGCGTTGTTGCTGTGAATGGAAGTGCATTCCATCGTCT
CATCTTCACCAGCCAACGATAATCTATGAACCAGGCGAAG

INTRON 1F-2/1G-1 (SEQ ID NO:120)

GTGAGATATATGCAAATTGAATGTTGTCCAGATGCGTTGTTTACATTTATATGCTTGGAA
TTGTCCTGAACGAATACAGTGGAAATAACCAAAAGCTGAAAAATAAAAAGATATATACTTC
ATTCTGAATTTGTGAGTATTGCTGACCCAAAACACGTTATCCATGTGACACTATATTT
GCCTTTCTGAATCTGAGACTGCGTTATGTTTCTAATAATCACGAAATATGGTATACAGGT
TGTGTATCTGTAGAATACCCAAGGCAGAAATTTAAAGGGTCACACCCTGTTTAATACAG

Fig. 4j

DOMAIN 1G-1 (1st part of domain g)

ATCACCATGACGACCATCAGTCGGGAAGCATAGCAGGATCCGGGGTCCGCAAGGACGTGA
ACACCTTGACTAAGGCTGAGACCGACAACCTGAGGGAGGCGCTGTGGGGTGTTCATGGCAG
ACCACGGTCCCAATGGCTTTCAAGCTATTGCTGCTTTCCATGGAAAACCAGCTTTGTGTG
CCATGCCTGATGGCCACAACACTCATGTTGTACTCACG

INTRON 1G-1/1G-2 (SEQ ID NO:121)

GTAAGTTTGTGTTGGTTAGTGTGTTGGTTGCATGTTTTGCCATATCGATAGTATCAGTGTGG
TAACATCTGGTTTCTAGTTCATTTCAGTTCACCTTATCAGAAGCTGTTTGCTCTCGTCTAC
AATAGTGACGTCTTTTCAGTTTTAGAACCGTGTACATCCGGGTATATTGGTCTCCAGCAA
CCCGTGCTTGTCGTGGGAGGCCACTGATGGGAACGGGTGGTCAGACTCGCTCACTTAGTT
GACACATGTCAATTGCGAAGATCGATGCTGAGGTTGTTAAACATTGGATTGTCTGGTCCA
GACTCGATTATTTACAGACAGCCGCCATGTACCTGGAATATTGCTGAGTGCGGCGTTAA
CAACAACTAGTCAGACTAATCTTTCACTGTTTTATAATGATGGCTCGAACCTAGCACTCA
TGTCCCAAGTTGGCGAACATCTGGAAGGGAATTTCAAATGAAAAGAACAATCTTTCACGT
CTATTGGTATCACGCTCCTGGAGAAGAACATGATGTTTACGGCGTTACTTCCTCTTACCT
GTTTTACTTGTTCCACGTTTCTTCATATTTAAAGAGTATTTGGGTATTAGAGCTTTGGT
GCTGTTACAATGCTACTCAACTGTTTCAGTGCGGGCGACCGCGCTTGTTTACACATTAAGT
TTTGTTTGTGTTGGTTGGTTTGTGTGTGTGTGTGTATGTGTGTGTGTGTGTGTGTGTGTA
TGTGTGTGTGTGTGTATCTATGTCTATGTGTCTGTGTCTGTGTGTCTGTCTATGTGTGTG
TGTGTCTGTGTCTATGTGTGTGTCTGCGTGTGTGTCTGTGTCCGTATGTGGCTGTGTCTA
TGTGTGTGTGTGTCTGTGTTTATGTGTGTATATGCGTGTGTGTCTGTGTCCGTATGTGGC
TGTGTCTATGTGTGTGACATGCAATACATGCTGTGATACTCACTAGCTGCGTCTATCGAC
CAG

DOMAIN 1G-2 (2nd part of domain g)

GCATGGCTACCTTCCCACACTGGCATCGCCTCTACACCAAGCAGATGGAGGATGCAATGA
GGGCGCATGGGTCTCATGTCGGCCTGCCCTACTGGGACTGGACTGCTGCCTTCACCCACC
TGCCAACACTGGTCACCGACACGGACAACAACCCCTTCCAACAT

INTRON 1G-2/1G-3 (SEQ ID NO:122)

GTAAGAGCGGGGTAGGGATGGGGTGGTAGGGGGTGGGTGTTCTATTACTTCCCGCTTCA
CTTGATGAAATGGATAACCTTGGCTGCATCCCAATTGCGTGATCGATTCTCTTTCGATT
CACTCGTGCGATTAGACTGCCTTATTTACTATAGTAGTTAGAATGTTGCTCAGTGCGCCG
TTAAACAATAACACAAAACCGCATTTGTTTTATATGGTCACTCTACTGTTTATCACG
TATATGTATGTTCCGACTCACTGGTTGGTGCGTACCATTCTACTGTCACTGAGAGCCA
ATGTTCTCAGATGTGTGAAATGTTTGAAAGCCGTTTCTACATAATATTGCAGGAATACCA
TTGTAGAATGTAGTCAAACAGGTAACAATCTGTTAGTGAGCCCAGTTCGAGGTTGCGTTG
TAGGGTGTAGTCCAACAGGTAGGCAGTCCATAAGCATAGTTTTTAAGCATTTTAGATCAT
CTATAATTAACCACATGGTTAGCCGCTATGTTTAGTTTAATCCAGTATAAGTTAGAAGTG
TTATATTTTCGAAGGGAAGTGAGTAAATCCTTATTCCTTGACTACCATTTAATAGATTTC
CAATGACTCCATTCAACTCCTAACTTTACATCACTGCTCTCTTCAACAG

DOMAIN 1G-3 (3rd part of domain g)

GGACACATTGATTATCTCAATGTCAGCACAACTCGATCTCCCCGAGACATGCTGTTCAAC
GACCCCGAGCATGGATCAGAGTCGTTCTTCTACAGACAAGTCCTCTTAGCTCTGGAACAA

Fig. 4k

ACTGATTTCTGCAAATTCGAAGTTCAGTTTGAGATAACCCACAATGCCATCCATTCTCTGG
ACAGGTGGCCACAGCCCCTACGGAATGTCCACTCTCGACTTCACTGCCTACGATCCTCTC
TTCTGGCTTCACCACTCCAACACCGACAGAATCTGGGCTGTCTGGCAAGCTTTGCAAGAA
TACAGAGGACTTCCATACAACCATGCCAATTGTGAGATCCAGGCAATGAAAACGCCCCTG
AGGCCTTTTCAGTGACGATATCAACCACAACCCAGTCACAAAGGCTAACGCGAAGCCATTA
GATGTGTTTCGAGTATAATCGGTTGAGCTTCCAGTACGACAACCTCATCTTCCATGGATAC
AGTATTCCGGAACCTTGATCGCGTGCTTGAAGAAAGAAAGGAGGAGGACAGAATATTTGCT
GCCTTCCTTCTCAGTGGAATCAAGCGTAGTGCTGATGTAGTGTTTCGACATATGCCAGCCA
GAACACGAATGTGTGTTTCGAGGGACTTTTTCGATTTTGGGAGGGGAGCTAGAAATGCCC
TGGTCCTTCGACAGACTGTTCCGCTATGATATCACCAAGGTGATGAAGCAGCTACACCTG
AGGCATGACTCTGACTTTACCTTCAGGGTGAAGATTGTCTGGCACCGACGACCACGAGCTT
CCTTCAGACAGTGTCAAAGCACCAACTATTGAATTTGAACCGGGCG

INTRON 1G-3/1H (SEQ ID NO:123)

GTGAGTACGACAGGCATTTCTAGTAAAAACCTACTTTTGGTAAAAGGTTTCGAGAAATCAC
TTGAAGCAACAACATGATTTTGTAAACGCCTATTACACGTGAACATGTCACACCCGGTGAT
GCCGTTTAATGGACATGCCTCTGTAAATGAAAGGGGTAAAGTACATGTGTATGGGGATGGG
ATGGGAGCCACCTGTCCCAATTTTCATAGGTCCCTAGGATCCCAGTTGCGTAGGAATCCCC
TGATTAATGCCTTGTGAATTCCTCCTGGAATTGTCTTGGCCCCAAATTTTACAAACCCGC
CCCGATATACCTTGGAAATAATTGGGCCTAAGGGTGGGGCTTTTAAGGACCAAGAACCCA
ACCTAAACCCCAACCCATTTTTTCCACCCATTCCAGGTTTTGTTTTACCAAATAAAAAG
GTTTCCACTTTGAGGAAACCCTTTAAGGGTCTTTTCAGGGCTTTTTTTCTTTTCTGGGA
ATTCCAATTCCGGGGGAACAAAATACATATATTTTCACAGACCTTTGGTCAAATTTATATA
ATTTCCGACTTCATGTCATAGGTTTGTCTTTCTTCTTACACAG

DOMAIN 1H

TGCACAGAGGCGGAAACCACGAAGATGAACACCATGATGACAGACTCGCAGATGTCCTGA
TCAGGAAAGAAGTTGACTTCCTCTCCCTGCAAGAGGCCAACGCAATTAAGGATGCACTGT
ACAAGCTCCAGAATGACGACAGTAAAGGGGGCTTTGAGGCCATAGCTGGCTATCACGGGT
ATCCTAATATGTGTCCAGAAAGAGGTACCGACAAGTATCCCTGCTGTGTCCACGGAATGC
CCGTGTTCCCCCACTGGCACCGCCTGCATACCATTGAGATGGAGAGAGCTCTGAAAAACC
ATGGCTCTCCAATGGGCATTCTTTACTGGGATTGGACAAAGAAGATGTGAGTCTTCCAT
CTTTCTTTGGAGATTCCAGCAACAACAACCCCTTTCTACAAATATTACATCCGGGGCGTGC
AGCACGAAACAACCAGGGACATTAATCAGAGACTCTTTAATCAAACCAAGTTTGGTGAAT
TTGATTACCTATATTACCTAACTCTGCAAGTCTTGAGGAAAACCTCGTACTGTGACTTTG
AAGTTCAGTATGAGATCCTCCATAACGCCGTCCACTCCTGGCTTGGAGGAACTGGAAAGT
ATTCCATGTCTACCCTGGAGCATTGCGCCTTTGACCCTGTCTTCATGATTACCACTCGA
GTTTGGATAGAATCTGGATCCTTTGGCAGAAGTTGCAAAAGATAAGAATGAAGCCTTACT
ACGCATTGGATTGTGCTGGCGACAGACTTATGAAAGACCCCTGCATCCCTTCAACTACG
AAACCGTTAATGAAGATGAATTCACCCGCATCAACTCTTTCCCAAGCATACTGTTTGACC
ACTACAGGTTCAACTATGAATACGATAACATGAGAATCAGGGGTCAGGACATACATGAAC
TTGAAGAGGTAATTCAGGAATTAAGAAACAAAGATCGCATATTTGCTGGTTTTGTTTTGT
CGGGCTTACGGATATCAGCTACAGTGAAAGTATTCATTTCATTTCGAAAACGATACAAGTC
ACGAAGAATATGCAGGAGAATTTGCAGTTTTGGGAGGTGAGAAGGAGATGCCGTGGGCAT
ATGAAAGAATGCTGAAATTGGACATCTCCGATGCTGTACACAAGCTTCACGTGAAAGATG
AAGACATCCGTTTTAGAGTGGTTGTTACTGCCTACAACGGTGACGTTGTTACCACCAGGC
TGTCTCAGCCATTCATCGTCCACCGTCCAGCCCATGTGGCTCACGACATCTTGGTAATCC
CAGTAGGTGCGGGCCATGACCTTCCGCCTAAAGTCGTAGTAAAGAGCGGCACCAAAGTCG

Fig. 4I

AGTTTACACCAATAGATTTCGTCGGTGAACAAAGCAATGGTGGAGCTGGGCAGCTATACTG
CTATGGCTAAATGCATCGTTCCCCCTTTCTCTTACCACGGCTTTGAACTGGACAAAGTCT
ACAGCGTCGATCACGGAGACTACTACATTGCTGCAGGTACCCACGCGTTGTGTGAGCAGA
ACCTCAGGCTCCACATCCACGTGGAACACGAGTAG

3'UTR

TTCACAG

INTRON 3'UTR (SEQ ID NO:124)

GTGAGGAGAAGGCCCCAGGCTAGCAGGGCAATGGATGAAGGAAATAGGGGCAAAGGGAAT
AGCAGTTACACCATCGACATTTCCAACCTCCTCAGAACTAATATATAGCCTTAATACAA
CCAGCCAAGACTCAACGGGGCAGCCGGGGTGGGGGGATTGTTGGTGGTTCGCTGTTTCAGACCA
GGGTGCAAAATATCAGTGCGCAAATCAACATGTTGCGTGTGACACACTGACACAGCAGTC
ATTGAACCTGCAGACCCATAACAGGAAAATGGGGCAGATACGATCAAAGACAGTGTAATA
TAGGGATAAGTAGGCATATGCAACCACCTGATGGAAATGAAAAGGGGTAAGTTTAAACCC
CGGCTACCAAAGGTCCAATGGTTCCTTAACCCAGCTTACGCTATCCCTCTAATTTTCAGTA
TTGAGCTGATTTCTGTGCGAGTTCATGTAACTGTATACTTTCTGTATTATTACAG

3'UTR

GTTGCTATGCCGACTGCGCTATATTGGTGAACGAGACGATGAGGACATCTCTGAAAGAGT
TCGCCAAGTGATGTGTAGGTCACGGAAGTATTGTTGAGCTAACAATATGATGATTTCAA
ATGACTTGGCGCTCTAGGACAAAGACATAATTCATCAGCACCCCTGTGCACCAACTCTTTG
TTTGCTGCAAACGTCTGACAAGCGACACGTCAATCAACAAGCTGTTCAAACCTCAAGTGGA
TGTAAGTAGAATCGTTGGGCCATCGTTCACAAAGTATTGACAGATGTCACACATGATGGC
GAGAAACACTTTTAGAACTTTTAATGACCTAGAGTGACTTGTAATATGTAAATATATTCT
TCAAAGACTCAGCTGAACTATTGTTGGATAACACATCAATTCCTCAACAAAATGCTTTA
TCTTCACATGGATGTATGTAATGTGGCCGGCAATAAAGTATATATATGTAT

Fig. 5a**Primary structure of the HtH1 protein****SIGNAL PEPTIDE**

LVQFLLVALVVGAGA

DOMAIN A

DNVVRKDVSHLTVDEVQALHGALHDVTASTGPLSFEDITSYHAAPASCDYKGRKIACCVHGMPSFP
FWHRAVYVQAERALLSKRKTVMGPYWDWTQTLTHLPSLVTEPIYIDSKGGKAQTNYWYRGEIAFIN
KKTARAVDDRLFEEKVEPGHYTHLMETVLDALQDEFCKFEIQFELAHNAIHVLVGGKFEYSMSNLE
YTSYDPIFFLHHSNVDRLFIAIWQRLQELRGKNPNAMDCAHELAHQQLQPFNRDSNPVQLTKDHSTP
ADLFDYKQLGYSYDSLNLNGMTPEQLKTELDERHSSKERAFAFRLSGFGGSANVVVYACVPDDDP
SDDYCEKAGDFFILGGQSEMPWRFYRPFYDVTEAVVHHLGVPLSGHYVVKTELFVNGTALSPDLL
PQPTVAYRPGK

DOMAIN B

GHLDPVHHRHDDDLIVRKNIDHLTREEEYELRMALERFQADTSVDGYQATVEYHGLPARCPRPDA
KVRFACCMHGMAFPHWHRLFVTQVEDALVRRGSPIGVPYWDWTKPMTHLPDLASNETYVDPYGH
HHNPFFNANISFEEGHHTSRMIDSKLFAPVAFGEHSHLFDGILYAFEQEDFCDFEIQFELVHNSI
HAWIGGSEDYSMATLHYTAFDPIFYLHHSNVDRLWAIWQALQIRRHKPYQAHCAQSVEQLPMKPFA
FPSPLNNEKTHSHSVPTDIYDYEEVLHYSYDDLTFGGMNLEIEEAIHLRQQHERVFAGFLLAGI
GTSALVDIFINKPGNQPLKAGDIAILGGAKEMPWAFDRLYKVEITDSLKTLSLDVDGDYEVTFKIH
DMHGNAALDLDLIPHAADVSEPAH

DOMAIN C

PTFEDEKHSLRIRKNVDSLTPREETNELRKALELLENDHTAGGFNQLGAFHGEPKWCPNPEAEHKVA
CCVHGMVFPWHRLALQAENALRKHGYSALPYWDWTRPLSQLPDLVSHEQYTDPSDHVVKHNP
WFNGHIDTVNQDTTRSVREDLYQQPEFGHFTDIAQQVLLALEQDDFCSEVQYEISHNFIHALVGG
TDAYGMASLRYTAYDPIFFLHHSNTDRIWAIWQSLQKYRGKPYNTANCAIESMRPLQPFGLSSAI
NPDRITREHAIPFDVFNRYRDNLHYVYDTLEFNGLSISQLDRELEKIKSHERVFAGFLLSGIKKSAL
VKFEVCTPPDNCHKAGEFYLLGDENEMAWAYDRLFKYDITQVLEANHLHFYDHLFIRYEVFDLKG
SLGTDLFHTANVVHDSGT

DOMAIN D

GTRDRDNYVEEVTGASHIRKNLNDLNTGEMESLRAAFLHIQDDGTYESIAQYHGKPGKCQLNDHNI
ACCVHGMPTFPQWHRLYVQVENALLNRGSGVAVPYWEWTAPIDHLPHFIDDATYFNSRQQRYDPN
PFRGKVTFENAVTTRDPQAGLFNSDYMENVLLALEQENYCDFEIQFELVHNALHSMMLGGKGQYS
MSSLDYSAFDPVFFLHHANTDRLWAIWQELQRFRELPYEEANCAINLMHQPLKPFSDPHENHDNVT
LKYSKPQDGFYQNHFGYKYDNLEFHLSIPSLDATLKQRRNHDRVFAGFLLHNIGTSADITIYIC
LPDGRGNDCSHEAGTFYILGGETEMPFIFDRLYKFEITKPLQQLGVKLHGGVFEELELEIKAYNGS
YLDPHTFDPTIIFEPGT

DOMAIN E

DTHILDHDHEEEILVRKNIIDLSRPRVSLVKALQRMKNDRSADGYQAIASFHALPPLCPNPSAAH
RYACCVHGMATFPQWHRLYTVQVQDALRRHGSLVGIPYWDWTKPVNELPELLSSATFYHPIRNINI
SNPFLGADIEFEGPGVHTERHINTERLFHSGDHDGYHNWFFETVLFALQEDYCDFEIQFEIAHNG

Fig. 5b

IHTWIGGS AVYGMGHLHYASYDPIFYIHHSQTDRIWAIWQELQKYRGLSGSEANCAIEHMRTPLKP
FSFGPPYNLNSHTQEYSKPEDTFDYKKFGYRYSLELEGRSISRIDELIQQRQEKDRTFAGFLLKG
FGTSASVSLQVCRVDHTCKDAGYFTILGSSAEMPWAFDRLYKYDITKTLHDMNLRHEDTFSIDVTI
TSYNGTVLSGDLIQTPSII FVPGR

DOMAIN F

HKLNSRKHTPNRVRHELSSLSSRDIA SLKAALTSLOHDNGTDGYQAIAAFHGVPAQCHEPSGREIA
CCIHGMATFPHWHRLYTQLQEQALRRHGSSVAVPYWDWTKPITELPHILTDGEYYDVWQNAVLANP
FARGYVKIKDAFTVRNVQESL FKMSSFGKHSLLFDQALLALEQTDYCDFEVQFEVMHNTIHYLVGG
RQTYAFSSLEYSSYDPIFFIHHSFVDKIWAVWQELQSRRLQFRTADCAVGLMGQAMRPFNKDFNH
NSFTKKHAVPNTVFDYEDLGYNNDNLEISGLNLNEIEALIAKRKSHARVFAGFLLFGLGTSADIHL
EICKTSENCHDAGVIFILGSSAEMHWAYNRLYKYDITEALQEFDINPEDVFHADEPFFLRLSVVAV
NGTVIPSSHLHQPTIIYEPGE

DOMAIN G

DHDDHQS GSIAGSGVRKDVNTLTKAETDNLREALWGMADHGPNGFQAIAAFHGKPALCPMPDGH
NYSCTHGMATFPHWHRLYTKQMEDAMRAHGSHVGLPYWDWTAFTHLPTLVTDTDNNPFQHGHI
YLVNSTTRSPRDMLENDPEHGSESFYRQVLLALEQTD FCKFEVQFEITHNAIHSWTGGHSPYGMS
TLDFTAYDPLFWLHHSNTDRIWAVWQALQEYRGLPNHANCEIQAMKTPLRPFSD DINHPVTKAN
AKPLDVFEYNRLSFQYDNLI FHGYSIPELDRVLEERKEEDRIFAFLLSGIKRSADVVDICQPEH
ECVFAGTFAILGGELEMPWSFDRLFRYDITKVMKQLHLRHDSDFTRVKIVGTDDHELPSDSVKAP
TIEFEPG

DOMAIN H

VHRGGNHEDEHDDRLADVLRKEVD FLSLQEANAIKDALYKLQND DSKGGFEAIAGYHGYPNMCP
ERGTDKYPCCVHGMPVFPHWHRLHTIQMERALKNHGSPMGI PYWDWTKKMSSLPSFFGDSSNNNPF
YKYYIRGVQHETTRDINQRLFNQTKFGEFDYLYLTQLVLEENSYCDFEVQYEILHNAVH SWLGGT
GKYSMSTLEHSADFDPVFMIIHSSLDRIWILWQKLQKIRMKPYALDCAGDRLMKDPLHPFNYETVN
EDEFTRINSFPSILFDHYRFNYEYDNMRIRGQDIHELEEV IQELRNKDRI FAGFVLSGLRISATVK
VFIHSKNDTSHEEYAGEFAVLGGEKEMPWAYERMLKLDISDAVHKLHVKDEDIRFRVVVTAYNGDV
VTTRLSQPFIVHRPAHVAHDILVIPVGAGHDLPPKVVKSGTKVEFTPIDSSVNKAMVELGSYTAM
AKCIVPPFSYHGFELDKVYSVDHGDYIIAAGTHALCEQNLRLHIHVEHE

DOMAIN 2A-1 (1st part of domain a)
[domain a, parts 1-4: SEQ ID NO:156]

INTRON 2A-1/2A-2 (SEQ ID NO:125)

DOMAIN 2A-2 (2nd part of domain a)

INTRON 2A-2/2A-3 (SEQ ID NO:126)

GTGAGTCACGTTCTCTGATGGTCACGAGTCACGTTCTCTGATGGTCACGAGTCACGTTCTCTGATG
GTCACGAGTCACGTTCTCTGATGGTCACGAGTCACATTCTCTGATGGTCACGAGTCACATTCTCTG
TTGAGTGAAGTCTCAGTACCATTTATTTCTCTTACCTTCTTCTAACCAGGGGTTTCAGCGTGGATC
GTCTGAGAAGTTAGCGCAAATCTATATTGAAGTCATTTTTCTATCATATAACCATCGTTATATCCA
CGTGCGAAAGTGTTCAATTAATTATTTTTATTTTTCAATTTATGAAGGTCTAAAAGAAAATATGTATTG
TTGGAAACTATATTCGAAGGTGAAGGCAACACGAGTGTATTAATATTCTCAATATCAATGTACGCT
CTGTCTAGCACCTGTTTCACCAGGAACACACCTTTAGCGTACCAAAATATCAGCTGATGATTTTCGA
AGCGGACTATACCCTCACCACCTTGTTTTGTGTGTGTATTTATGTGTGCATGTGTGTGCGTGCGTGC
GTGTGTGTGTGTGTCCTACGTATGTTGATATTTTGTCTGACTGTATATGTTTCGTGCTTACCATTG
AAG

Fig. 6b

DOMAIN 2A-3 (3rd part of domain a)

GTACTCCATGTCTCATCTCGAGTACACCTCCTACGACCCCCCTCTTCTTCCATCACTCCAACAC
CGACCGCATCTTTCGCCATCTGGCAACGTCTTCAGGTACTCAGAGGAAAGGACCCCAACACCGCCGA
CTGCGCACACAACCTCATCCATGAGCCCATGGAACCGTTCCGTCGGGACTCGAACCCCTCTTGACCT
CACCAGGGAAAACCTCCAACCAATTGACAGCTTTGATTATGCCACCTTGGCTACCA

INTRON 2A-3/2A-4 (SEQ ID NO:127)

GTATGTATGATTCTAATAATGAATGTTTTTACCTCCGGTTTAAACAATATTTTAGTATTACGAAAG
GAGAAGTACCTCGAGAGGTCTAGGTCTCAGATGTTTAGAAACCCATGAAGACAGGTATGCTTCTGA
AAAACAAAGTAACATCATGAGGCTAAAGTTCAGATTCAAACCATCGTAGTTCGAATCCAGCATGCA
AAGGGCCCTAACCTGTAGATGGCGCTGCTTGAAACAGAGTAGTCTGTTTCAGGGTCAGTACTGTCC
CCACAAACATCATAGTCAGGGTCAGTACTGTCCCCACAAACATCATAGTCAGGGTCAGTACTGTCC
CCACAAACATCACAGTCAGGGTTAATTTTGGATTTCGGTTTTCGAATGCGAAGAAGACAGTCACGCCC
TGACACTGGACCGAGGTTGCCGAGAAAGCTCGTGATATTGCTGGAATACTGCCCAGTAAAACCATC
ATTTATTTTtaggctATTTATTACGAAAAATAATAATATGTATAGAAATGCATATGATCGCTGTTTG
AATGTAAATTTAGAATGGGTTTGGGAGTGTTCACTATTTTTTTCATCAAAATTTTCATGTATTTTAA
CCGATCGACGCTGAAGACAAACTACCGTTAATCAGGCAGTTCATTTCATATCTGATAGGGAATATTG
GTTGTTAACCAACGCTACATTGTGTCCAG

DOMAIN 2A-4 (4th part of domain a)

GTATGATGACTTGACCCTGAACGGTATGACCCCAGAGGAATTGAACTCATATCTGCATGAACGGTC
AGGCAAGGAGGGGGTGTTCGCAAGCTTCCGACTCTCAGGTTTTGGCGGCTCTGCTAACGTTGTTGT
CTACGCATGCCGTCTGCCCACGATGAAATGGCTGTGATCAGTGCGACAAAGCCGGCGACTTCTT
TGTGTTGGGCGGACCCACCGAGATGCCCTGGAGGTTTTACAGAGCATTCCTACTTCGACGTCACCGA
CAGCATCGACAACATCGACAAGGACCGCCACGGCCACTATTATGTAAAGGCGGAATTATTCAGTGT
AAATGGAAGTGCCTACCGAATGATCTCCTGCCTCAACCCACCATCTCACACAGGCCAGCCCGCGG
ACACGTTGATG

INTRON 2A-4/2B (SEQ ID NO:128)

GTAAATGGCCATTGTATACATGCATTCATTTGGACTTTGAGTGAGTGAGTGAGTGCGTATTCAGTA
AGTGAGAGTGTGAGTGGGTATTAGGTCTGTGAGTGGGTGGTGAGTGAGTGAGTAAAGAGTGG
GTTGGTGAGAAAGTGAGTGAGTCACTTGGTGGGTGCGTTAGTGGAAGCGTGATTGAGTGAGTGGA
GGTAGGTGAGTGAGTGAATTGGTGGGGGGGTGAGTGAGGTTAACGCTGTTCTGCTGTTCAATCACA
CCACATGTTGCCAGCTTACTGTGCAGGACGAATCCAGGGTTGTGTTAAATTTTATATGTTTATATA
TAACGATGGACGTGTCTGGATGTGGCGAATGTGTCAAGAGAATTATGCGGCTTTGTGCTGCTCCGC
GTATTTTATTGCACGCGCGTTGGTACGCGGTTGATAAAGTAGTTCAAACATTTCCCAGCCATCTTT
GTCTGTTGTGAAAACCTACTCCAGGACCATCCATTTCAATATGTGTCTGCGTTCATGGAGTTATAC
ATGTTAAACTGTAGAGCGCAGATGAGCACACTTGAGCATTTCTTCAGTAAATCAGAATGTGTATAT
TTCAAAATTTACCAAATGCAATATCATCAAGCAAATTATGCAGCTCTATAGTAACATCGGAGTCAA
TGGTCCAGTGTGCCCTCGGCTGCCATTCCGACCTCCCTGGCCAGAATACACCCCGGTGAGGATCAG
TTATCCGTCAGAAGGCACGGTGCGGAATGAAAACATAAACACATAGTCGCTTAGTAGTATGCTGAT
TTAGGCACGCAAAATCCGAATGTGAATTACTGTGAATTGCATTACCTGTTACAG

Fig. 6c

DOMAIN 2B

AGGCCCCAGCTCCCTCCTCGGATGCTCACCTCGCCGTCAGGAAGGATATCAACCATCTGACACGCG
AGGAGGTGTACGAGCTGCGCAGAGCTATGGAGAGATTCCAGGCCGACACATCCGTTGATGGGTACC
AGGCTACGGTTGAGTATCACGGCTTACCTGCTCGATGTCCATTCCTCCGAGGCCACAAATAGGTTCG
CCTGTTGCATCCACGGCATGGCGACATTCCCTCATTGGCACAGACTGTTCTGTTACCCAGGTGGAAG
ATGCACTGATCAGGCGAGGATCCCTATAGGGGTCCCTACTGGGACTGGACTCAGCCTATGGCAC
ATCTCCCAGGACTTGCAGACAACGCCACCTATAGAGATCCCATCAGCGGAGACAGCAGACACAACC
CGTTCCACGATGTTGAAGTTGCCTTTGAAAATGGGCGTACAGAACGTCACCCAGATAGTAGATTGT
TTGAACAACCTCTATTTGGCAAACATACGCGTCTCTTCGACAGTATAGTCTATGCTTTTGAGCAGG
AGGACTTCTGCGATTTTGAAGTTCAATTTGAGATGACCCATAATAATATTCACGCCTGGATTGGTG
GCGGCGGGAAGTATTCCATGTCTTCTCTACACTACACAGCCTTCGACCCTATCTCCTACCTTCATC
ACTCCAACACTGACCGTCTCTGGGCAATTTGGCAAGCGTTGCAGATACGAAGAAACAAACCGTATA
AGGCTCATTGTGCTTGGTCTGAGGAACGCCAGCCTCTCAAACCTTTCGCCTTCAGTTCCTCCACTGA
ACAACAACGAAAAACCTACGAAACTCGGTGCCACCAACGTTTACGACTACGAAGGAGTCCTTG
GCTATACTTATGATGACCTCAACTTCGGGGGCATGGACCTGGGTGAGCTTGAGGAATACATCCAGA
GGCAGAGACAGAGAGACAGGACCTTTGCTGGCTTCTTTCTGTACATATTGGTACATCAGCGAATG
TTGAAATCATTATAGACCATGGGACTCTTCATACCTCCGTGGGCACGTTTGCTGTTCTTGCGGAG
AGAAGGAGATGAAATGGGGATTTGACCGTTTGTACAAATATGAGATTACAGATGAACTGAGGCAAC
TTAATCTCCGTGCTGATGATGGTTTCAGCATCTCTGTAAAGTAACTGATGTTGATGGCAGTGAGC
TGTCCTCTGAACTCATCCCATCTGCTGCTATCATCTTCGAACGAAGCCATA

INTRON 2B/2C (SEQ ID NO:129)

GTAAGTAGCTACCTGTTTATTCAATTTTTTCGCTTTGCCAATCAATTCATTTCAGCTTGAAATTCAA
TAATTGTGTTTTGCATGGCTGAAAACCAATTTGAACTCTTTTCTTTTCTCAGGTGCAACTCAAATA
AATAATCACTAATTGTTATGCACGCGGGTAGGGCATACTATATCCACATCGGTTCATCTCAA
ATGCAACAAATTGTCTTATTTCCGTTGGGACAAGCAAACCCCTTTCCTGTAATCTTGCCTTTGG
CATCCACTGGAATTAATGTTGACTGGTAATTGATACTGGCTCTCTTCTTGATAGAGTTAATATCT
ATAGTTTGTAAATCTTTATGATTTTGCTATTTATATTTGCACAGCATGCTATAGACACCCTAGACT
ATTGTATAGCCACTTGATTTGTTTTCCATTTATTATTTATAACAGAACATGGCTTGTAATTTTAA
TTTACCTTCCAG

DOMAIN 2C

TTGACCATCAGGACCCTCATCAGGACACAATCATCAGGAAAAATGTTGATAATCTTACACCCGAGG
AAATTAATTCTCTGAGGAGGGCAATGGCAGACCTTCAATCAGACAAAACCGCCGGTGGATTCCAGC
AAATTGCTGCTTTTCACGGGGAACCCAAATGGTGCCCAAGTCCCGATGCTGAGAAGAAGTTCTCCT
GCTGTGTCCATGGAATGGCTGTCTTCCCTCACTGGCACAGACTCCTGACCGTGCAAGGCGAGAATG
CCCTGAGAAAGCATGGATGTCTCGGAGCTCTCCCTACTGGGACTGGACTCGGCCCTGTCTCACC
TACCTGATTTGGTAAGTCAGCAGAACTACACCGATGCCATATCCACCGTGGAAGCCCGAAACCCCT
GGTACAGCGGCCATATTGATACAGTTGGTGTGACACAACAAGAAGCGTCCGTCAAGAAGTGTATG
AAGCTCCCGGATTTGGTCATTATACTGGGGTCGCTAAGCAAGTGCTTCTGGCTTTGGAGCAGGATG
ACTTCTGTGATTTTGAAGTCCAGTTTGAGATAGCTCACAATTTTCATCCACGCTCTTGTGCGCGGAA
GCGAGCCATATGGTATGGCGTCACTCCGTTACACTACTTATGATCCAATTTTCTACCTCCATCATT
CTAACACTGACAGACTCTGGGCTATATGGCAGGCTCTACAAAAGTACAGGGGCAAACCTTACAATT
CCGCCAACTGTGCCATTGCTTCTATGAGAAAACCCCTACAGCCCTTTGGTCTGACTGATGAGATCA
ACCCGGATGATGAGACAAGACAGCATGCTGTTCCCTTTCAGTGTCTTTGATTACAAGAACAACCTTCA
ATTATGAATATGACACCCTTGACTTCAACGGACTATCAATCTCCAGCTGGACCGTGAAGTGTGAC
GGAGAAAGTCTCATGACAGAGTATTTGCCGGATTTTGTGCTGCATGGTATTTCAGCAGTCTGCACTAG

Fig. 6d

TTAAATTCTTTGTCTGCAAATCAGATGATGACTGTGACCACTATGCTGGTGAATTCTACATCCTTG
GTGATGAAGCTGAAATGCCATGGGGCTATGATCGTCTTTACAAATATGAGATCACTGAGCAGCTCA
ATGCCCTGGATCTACACATCGGAGATAGATTCTTCATCAGATACGAAGCGTTTGATCTTCATGGTA
CAAGTCTTGGAAGCAACATCTTCCCCAAACCTTCTGTCATACATGACGAAGGGGCAG

INTRON 2C/2D (SEQ ID NO:130)

GTGAGAACATTGATAATAGTTCAAATGAAGTATATCCGATTCAAGCTGTCGATACAAGATGAGATA
CATAATCACAATGTTTGTATTAGATATCTCTCTTAATTTAATGCCGCTTTTATCAATATTCGAGCA
ATCCTTCAGCAACATACACCAGCAAATGTTTCATCAACAGACTATATTATTTAATATTTTAAAAAT
CCTTCTCTGTTGTTATAAATACTTAAAGTATCGAATTCCTTGAATGCGTCTTCTCTGCAGCATATA
GTTAAGTTGTTGTGTTTCTCTGTCAG

DOMAIN 2D

GTCACCATCAGGCTGACGAGTACGACGAAGTTGTAAGTCTGCAAGCCACATCAGAAAGAATTTAA
AAGATCTGTCAAAGGGAGAAGTAGAGAGCCTAAGGTCTGCCTTCCTGCAACTTCAGAACGACGGAG
TCTATGAGAATATTGCCAAATTCACGGCAAGCCTGGGTTGTGTGATGATAACGGTTCGCAAGGTTG
CCTGTTGTGTCCATGGAATGCCACCTTCCCCCAGTGGCACAGACTCTATGTCTCCAGGTGGAGA
ATGCTTTGCTGGAGAGAGGATCTGCCGTCTCTGTGCCATACTGGGACTGGACTGAAACATTTACAG
AGCTGCCATCTTTGATTGCTGAGGCTACCTATTTCAATTCCCGTCAACAAACGTTTGACCCTAATC
CTTTCTTCAGAGGTAAAATCAGTTTTTGAGAATGCTGTTACAACACGTGATCCCCAGCCTGAGCTGT
ACGTTAACAGGTACTACTACCAAAACGTCATGTTGGCTTTTGAACAGGACAACACTACTGCGACTTCG
AGATACAGTTTGAGATGGTTCACAATGTTCTCCATGCTTGGCTTGGTGGAAGAGCTACTTATTCTA
TTTCTTCTCTTGATTATTCTGCATTTCGACCCTGTGTTTTTCTTCCCATGCGAACACAGATAGAT
TGTGGGCCATCTGGCAGGAGCTGCAGAGGTACAGGAAGAAGCCATACAATGAAGCGGATTGTGCCA
TTAACCTAATGCGCAAACCTCTACATCCCTTCGACAACAGTGATCTCAATCATGATCCTGTAACCT
TTAAATACTCAAAACCCACTGATGGCTTTGACTACCAGAACAACCTTTGGATACAAGTATGACAACC
TTGAGTTCAATCATTTTCAGTATTCCCAGGCTTGAAGAAATCATTCGTATTAGACAACGTCAAGATC
GTGTGTTTGCAGGATTCCTCCTTCACAACATTGGGACATCCGCAACTGTTGAGATATTTCGTCTGTG
TCCCTACCACCAGCGGTGAGCAAACTGTGAAAACAAAGCCGGAACATTTGCCGTACTCGGAGGAG
AAACAGAGATGGCGTTTCATTTTGACAGACTCTACAGGTTTGACATCAGTGAAACACTGAGGGACC
TCGGCATAACAGCTGGACAGCCATGACTTTGACCTCAGCATCAAGATTCAAGGAGTAAATGGATCCT
ACCTTGATCCACACATCCTGCCAGAGCCATCCTTGATTTTTTGTGCCTGGTTCAA

INTRON 2D/2E (SEQ ID NO:131)

GTAAGAAAGTTTCACTGTCTAAATCTTTTTTTATGATAGAGGGTAGAGAAGTGGAGACAATGTGAC
AATATATTGAATAAAGTTGTTTAAAATTTATAACTCTCATAAGTTCATATTATGCTGAAGCTGTAG
CCATCTATAACTGTGTAACATGAAATGTTAAGACATTAACCTAAATACTTCAGCTGATAACAAAAC
AATGTTAATACATACGTCAATGTAACATTTTCTTATCTTTAGGTTATAGCATAAAACACTTCAGAGA
TACAGTGACGAAAACCTCTATTTAAATATTTTCAG

DOMAIN 2E

GTTCTTTCTGCGTCCTGATGGGCATTTCAGATGACATCCTTGTGAGAAAAGAAGTGAACAGCCTGA
CAACCAGGGAGACTGCATCTCTGATCCATGCTCTGAAAAGTATGCAGGAAGACCATTACCTGATG
GGTTCCAAGCCATTGCCTCTTTCCATGCCCTGCCACCACTCTGCCCTTCACCATCTGCAACTCACC
GTTATGCTTGCTGTGTCCACGGCATGGCTACATTTCCCCAGTGGCACAGACTGTACACTGTACAGT

Fig. 6e

TCCAGGATGCACTGAGGAGACATGGAGCTGCAGTAGGTGTACCGTATTGGGATTGGCTGCGACCGC
AGTCTCACCTACCAGAGCTTGTCCACCATGGAGACATACCATGATATTTGGAGTAACAGAGATTTCC
CCAATCCTTTCTACCAAGCCAATATTGAGTTTGAAGGAGAAAACATTACAACAGAGAGAGAAGTCA
TTGCAGACAAACTTTTTGTCAAAGGTGGACACGTTTTTTGATAACTGGTTCTTCAAACAAGCCATCC
TAGCGCTTGAGCAGGAAACTACTGTGACTTTGAGATTTCAGTTTGAAATTCTTCACAACGGCGTTC
ACACGTGGGTGCGAGGCAGTCGTACCCACTCTATCGGACATCTCCATTACGCATCCTACGACCCTC
TTTTCTACCTCCACCATTCCCAGACAGACCGTATTTGGGCAATCTGGCAAGAACTCCAGGAACAGA
GAGGGCTCTCAGGTGATGAGGCTCACTGTGCTCTCGAGCAAATGAGAGAACCATTGAAGCCTTTCA
GCTTCGCGCTCCTTATAACTTGAATCAGCTAACACAGGATTTCTCCCGACCCGAGGACACCTTCG
ACTACAGGAAGTTTGGTTATGAATATGACAATTTAGAATTCCTAGGAATGTCAGTTGCTGAACTGG
ATCAATACATTATTGAACATCAAGAAAATGATAGAGTATTCGCTGGGTTCTGTTGAGTGGATTTCG
GAGGTTCCGCATCAGTTAATTTCCAGGTTTGTAGAGCTGATTCACATGTCAGGATGCTGGGTACT
TCACCGTTCTTGGTGGCAGTGCTGAGATGGCGTGGGCATTTGACAGGCTATACAAATATGACATTA
CTGAAACTCTGGAGAAAATGCACCTTCGATATGATGATGACTTCACAATCTCTGTCTGAGTCTGACCG
CCAACAACGGAAGTGTCTGAGCAGCAGTCTAATCCCAACACCGAGTGTTCATATTCAGCGGGGAC
ATC

INTRON 2E/2F-1 (SEQ ID NO:132)

GTAAGTAGTAACTGCTCAGATTGTTTTTCATAATTACTCCACTATTAAGTAAAAAGTACTAGTAAT
TCAATAGTACTGTTTCACAGAGAAATGTAACACAATAGACCACAGAGTCCATTTGTTAAACGCCTTT
GGCTTGGTAAAGTCTGAGATTTTGGTGACTGATGGAAAGCTAAAATATATTTTGACAG

DOMAIN 2F-1 (1st part of domain f)

GTGACATAAATACCAAGAGCATGTCAGCGAACCGTGTTCCGCCGTGAGCTGAGCGATCTGTCTGCGA
GGGACCCGTCTAGTCTCAAGTCTGCTCTGCGAGACCTACAGGAGGATGATGGCCCCAACGGATACC
AGGCTCTTGCAGCCTTCCATGGGCTACCAGCAGGCTGCCATGATAGCCAGGGAAATGAG

INTRON 2F-1/2F-2 (SEQ ID NO:133)

GTATATTTAAGTATTTTATCTTACGCATGACCCTGACCCTATTTATTTTTTTTTTAATCCTCGGATT
TGTTTAAATCCTGTTACCAGCGAAGGTCCGGGTAGAATTGATCTTCAGTCAACTATTCTTGTCTGTA
GGACTAACGAGTTGTCTGGCTTGCTTACTCGGTTGACACGTGTCAACGGATCCCAATTGCAATTAG
ATCGATGCTCATGCTGTTGATCCCTGGATTGCCTGGTCCGGACTCCACATACCGCCGCATATTGC
TGGTATATTGTCGAATGCGACGCTAAACAGCAAGCCAACCAACAATACTGAGACCTGGTGGTACAT
GTCAGTTCTCTATTGCTGGGGTTCCAAACATAGCCATCAGTTGAAATATTTTCATACATAGAAGAAT
ACCTCTGAATATGATGATGAAACATTTACTTAGACTTGCTGTGAGCCCCAGGCCAAAATGCACTGT
AAAAATACACTGACAGAGGATTAGGCATTCTTGGGAGTACTGTATAGTTAGTTGCATACATATTAG
CGTTCCCTCACTAAAACGAATCTCTGAATGCTATCAATTAAAGATCATGATGCTTTGATTGTGTCT
ACTGTATTTAAATGGTGTAAAGATTTGCAATTACAATATACACAAACACGTTTCCTGCATCTCGG
AGAATGCAATCTTTCGTTGTACGCGTCTGTTTTTCATATTTTATGCATGTAGTTTGCCTACTTAG
CGTCCAATAAATCCATTACAAAATCACAAAACAAACGATTTTAGGAATGTGACTGTAGCTGCAA
CGAATATACCTGATCCTTTCTTGTTCAG

DOMAIN 2F-2 (2nd part of domain f)

ATCGCATGTTGCATTCACGGTATGCCGACCTTCCCCAGTGGCACAGACTGTACACCCTGCAGTTG
GAGATGGCTCTGAGGAGACATGGATCATCTGTGCGCATCCCCTACTGGGACTGGACAAAGCCTATC

Fig. 6f

TCCGAAC TCCCCTCGCTCTTCACCAGCCCTGAGTATTATGACCCATGGCATGATGCTGTGGTAAAC
AACCCATTCTCCAAAGGTTTTGTCAAATTTGCAAATACCTACACAGTAAGAGACCCACAGGAGATG
CTGTTCCAGCTTTTGTGAACATGGAGAGTCAATCCTCTATGAGCAAACCTCTTCTTGCTCTAGAGCAA
ACCGACTACTGTGATTTTGTAGGTACAGTTTGTAGGTCTCCATAACGTGATCCACTACCTTGTGGC
GGACGTACAGACCTACGCATTGTCTTCTCTGCATTATGCATCCTACGACCCATTCTTCTTTATACAC
CATTCTTTTGTGGATAAGATGTGGGTAGTATGGCAAGCTCTTCAAAGAGGAGGAACTTCCATAC
AAGCGAGCTGACTGTGCTGTCAACCTAATGACTAAACCAATGAGGCCATTTGACTCCGATATGAAT
CAGAACCCATTCACAAAGATGCACGCAGTTCCCAACACACTCTATGACTACGAGACACTGTACTAC
AGCTACGATAATCTCGAAATAGGTGGCAGGAATCTCGACCAGCTTCAGGCTGAAATTGACAGAAGC
AGAAGCCACGATCGCGTTTTTGTCTGGATTCTTGCTTCGTGGAATCGGAACCTTCTGCTGATGTCAGG
TTTTGGATTTGTAGAAATGAAATGACTGCCACAGGGGTGGAATAATTTTCATCTTAGGTGGAGCC
AAGGAAATGCCATGGTCATTTGACAGAACTTCAAGTTTGATATCACCCTGTACTCGAGAAAGCT
GGCATTAGCCAGAGGACGTGTTTGATGCTGAGGAGCCATTTTATATCAAGGTTGAGATCCATGCT
GTTAAACAAGACCATGATACCATCGTCTGTGATCCAGCCCCAACTATCATCTATTCTCTGGGGAA
G

INTRON 2F-2/2G-1 (SEQ ID NO:134)

GTGAGAGAACCAGTAATAGCTACTGTCTACAAAGAATGTGTTTCAATTTAAAGACCTGACTGTAGGCC
GATGGCTGCTGTCATCTCCTCCGCCTCCTCCTCCTGTTCCCTCCTCCGAAGGGGTCAGCTTCAGGTT
CTCTTGCCAATATGCCAAGCAGACCTCCTGAGCAGGCAGTATATATACGTAAGGGAAGCAAGTATG
GACCATCGCGCGGCATGTAGAGATACAATGATCAGCTGTCTGCTGTTCCACTCCTGTCAGACAATG
AGATAAACATGAATACAGTATTACTCAGCAGCGTTCCAATTTTCAACCCTCGTATTTATTAATAA
AGGAATTTTAAATATATTTTTCTCCTTGTTGAAATATTTTAGTAACTGTTAATCGATATAGAGTGG
AGTAGTGACGCTTTATTTTCGGTTCATTCTCGAAACAAAAATATAATAGTCCACTGAACTCTCTTAA
ATTGTTTTTACAACCTTCAACTGCCACAGACGTAATCCCTCACGTTATTTTGAGCTGACAACGTGT
TGAATTGAGTGTGTTCCGAATTCTAAATAAGCATGTATATATTTACGTCTCATGCAAGTAATATAT
GTTTAACTGATGACGTCACCTGGTGACCACTGATTTAGTTTCTTTGTCATAATTGCAGTTTCTGTT
GTCACGGGGACGGTGGGGAAGCCAGGTTCTCCTGTCACGCTGAATATCCCGTTCGAATCCCCCAC
ATGGGTACAAAGTGTGATGCCTATTTCTGGTGTCCCCCACCCTGATATTGCTGGAATAAGTGGCTT
AATACCATATACACTCACTCTATTGTCACTACTGCCACCGGCTCACACCTCTGATGCTTCTGTT
CTATCCAG

DOMAIN 2G-1 (1st part of domain g)

GTCGCGCTGCTGACAGTGCACACTCAGCCAACATTGCTGGCTCTGGGGTGAGGAAGGACGTCACGA
CCCTCACTGTGTCTGAGACCGAGAACCTAAGACAGGCTCTTCAAGGTGTCATCGATGATACTGGTC
CCAATGGTTACCAAGCAATAGCATCCTTCCACGGAAGTCCTCCAATGTGCGAGATGAACGGCCGCA
AGGTTGCCTGTTGTGCTCACG

INTRON 2G-1/2G-2 (SEQ ID NO:135)

GTAATTAATGGATGTGAAGTCAATGTCCGAGGGTATAATAAGGATTTAAATACTTCAGTCGTGTAA
TACTGTATGACATGTGTATTGGATGGTGTAGGTATTACAGGTATAAGGCCAGTGTGTGTTGGGAC
GGTTACTTTTCTGCACTAGTAATAAGCATTGTATTTAGCTAGCTTTTATCATATAACTTTAGTTTC
ATGGTTTGTGGCAATTGAAATCGAAATTTTCTTTCAATTTCAAGGTTATCGCACTCGTGTGTTAGAA
TAGTTACTATGCTGCATTGAGAATAACACTATAGTAATAAAGCATATCATACAGTAAGAATAACAC
TATAGTAATAAAGTATATCATACAGTAAGAATGTCATTGTATGATAAATAGGTTATCACACTCGTG
TGTTTTAGAAATGGTTACTATCCCAGGAATAACCACTATGTATTACATGTATATTGGGCAGTGTAAG
TAGTAGCATTGTATATTAAATCAGTATATCGTGCTTCAAACACCAGGATATATGGGGTATACAGT

Fig. 6g

GGGCAGTGTAAGTAGCAACATTGTATATTAAATCAGTATATCGTACTTCAAAACACCAGGATTATG
GGGTATACAGTGGGCAGTGTAAGTAGTAGCATTGTATATTAAATCAGTATATCGTACTTCAAAACA
CCAGGATATAATTCAGTATATCGTGCTTCAAAACACCAGGATATAATTCAGTATATCGTGCTTCAA
AACACCAGGATATATGGGATATACAGTGCGGGTTTGCATACAACCTCCACCCTTTACAG

DOMAIN 2G-2 (2nd part of domain g)

GTATGGCCTCCTTCCCACACTGGCACAGACTGTATGTGAAGCAGATGGAAGACGCCCTGGCTGACC
ACGGATCACATATCGGCATCCCTTACTGGGACTGGACAACCTGCCTTCACAGAGTTACCCGCCCTTG
TCACAGACTCCGAGAACAATCCCTTCCATGAG

INTRON 2G-2/2G-3 (SEQ ID NO:136)

GTCAGTTTAGTCTCCTGTCTGAGCTAACGATACCAATTTCTTATTTTCGAGAACCACGATGACGAG
AAAACAAGCAATATAGATATAGATGCAGTATAGATCAAGTTAATGAATTCATTGCTATATGTTTGC
TTGTAATAAACTTTAAGAAAACGAGAGCATGCACACAAATGAAACAAACAATTATGTGTTTGATAG
GAATATGATATATGTATTTGGGGGCTGACGTGAGCAGGGTTGAAGGGACAGTTTACATTGTGAGTA
ACACTGGGAGTATTCTTTGATCCACAATATATAGTTTCATTGTGTTGAGCAGTTACAACCTAACATT
ATATCATACATTACGTCGTAACATGCTTCTTTTGTCTCTTCTGCCAG

DOMAIN G-3 (3rd part of domain g)

GGTCGCATTGATCATCTCGGTGTAACCACGTCACGTTCCCCCAGAGACATGCTGTTTAAACGACCCA
GAGCAAGGATCAGAGTCGTTCTTCTATAGACAAGTCCTCCTGGCTTTGGAGCAGACTGACTACTGC
CAGTTTGAAGTCCAGTTTGAAGCTGACCCACAACGCCATTCACTCCTGGACAGGTGGACGTAGCCCT
TACGGAATGTCGACCCTCGAGTTCACAGCCTACGATCCTCTCTTCTGGCTTACCACTCCAACACC
GACAGAATCTGGGCTGTCTGGCAAGCACTGCAGAAATACCGAGGACTCCCATAACAACGAAGCACAC
TGTGAAATCCAGGTTCTGAAACAGCCCTTGAGGCCATTCAACGATGACATCAACCACAATCCAATC
ACCAAGACTAATGCCAGGCCTATCGATTCAATTTGATTATGAGAGGTTTAACTATCAGTATGACACC
CTTAGCTTCCATGGTAAGAGCATCCCTGAACTGAATGACCTGCTCGAGGAAAGAAAAAGAGAAGAG
AGAACATTTGCTGCCTTCTTCTCGTGGAATCGGTTGCAGTGCTGATGTCGTCTTTGACATCTGC
CGCCCCAATGGTGACTGTGTCTTTGCAGGAACCTTTGCTGTGCTGGGAGGGGAGCTAGAAATGCCT
TGGTCCTTCGACAGACTGTTCCGCTATGACATCACCAGAGTCATGAATCAGCTCCATCTCCAGTAT
GATTGAGATTTGAGTTTCAAGGTTGAAGCTTGTGCAACCAATGGCACTGAGCTTTCATCAGACCTC
CTCAAGTCACCAACAATTGAACATGAACTTGGAG

INTRON 2G-3/2H (SEQ ID NO:137)

GTATGTTATCTTATTATCAAATGTGTAATCAGATACTGGAGACGTTTTTCATATTAACCTGGTCAGC
ATTAGTTGATGATTTTGGTGCGATATTGACGACAAGGAGTTAAGCATTAACACGTTCAACACATCT
TTAATCTGATATGAGAAGGGAATAAATTGATCCAGTATTGATGATTGAAGTTAGATTAAACAGTGAA
AGATATACCAGTTTTGATAATCGTATAAAACAGTAGCAGAATTGTATCGTGAAAACCTAAATGTGGG
AAGGCGAACGCCAAGCAGATTTTAGATTACGATCGTGTGCTAGAATAATTCACAATAACCCAGACG
TCGGAATGTGGTTGTCTATGGCAATAGTTACGATTAATTGCTAACATGCACGATTTACCTATTTTC
AG

DOMAIN 2H

CCCACAGAGGACCAGTTGAAGAAACAGAAGTCACTCACCAAATACTGACGGCAATGCACACTTCC
ATCGTAAGGAAGTTGATTCGCTGTCCCTGGATGAAGCAAACAACCTGAAGAATGCCCTTTACAAGC

Fig. 6h

TACAGAACGACCACAGTCTAACAGGATACGAAGCAATCTCTGGTTACCATGGATACCCGAATCTGT
GTCCGGAAGAAGGCGATGACAAATACCCCTGCTGCGTCCACGGAATGGCCATCTTCCCCACTGGC
ACAGACTCTTGACCATCCAACCTGGAAAGAGCTCTCGAGCACAATGGTGCACCTGCTTGGTGTTCCTT
ACTGGGACTGGACCAAGGACCTGTCGTCACCTGCCGGCGTTCTTCTCCGACTCCAGCAACAACAATC
CCTACTTCAAGTACCACATCGCAGGTGTTGGTCACGACACCGTCAGAGAGCCAACTAGTCTTATAT
ATAACCAGCCCCAAATCCATGGTTATGATTATCTCTATTACCTAGCATTGACCACGCTTGAAGAAA
ACAATTACTGTGACTTTGAGGTTTCACTATGAGATCCTCCACAACGCCGTCCACTCCTGGCTTGGAG
GATCCCAGAAGTATTCCATGTCTACCCTGGAGTATTCGGCCTTTGACCCTGTCTTTATGATCCTTC
ACTCGGGTCTAGACAGACTTTGGATCATCTGGCAAGAACTTCAGAAGATCAGGAGAAAGCCCTACA
ACTTCGCTAAATGTGCTTATCATATGATGGAAGAGCCACTGGCGCCCTTCAGCTATCCATCTATCA
ACCAGGACGAGTTCACCCGTGCCAACTCCAAGCCTTCTACAGTTTTTGGACAGCCATAAGTTCGGCT
ACCATTACGATAACCTGAATGTTAGAGGTCACAGCATCCAAGAAGTCAACACAATCATCAATGACT
TGAGAAACACAGACAGAATCTACGCAGGATTTGTTTTGTTCAGGCATCGGTACGTCTGCTAGTGTCA
AGATCTATCTCCGAACAGATGACAATGACGAAGAAGTTGGAACCTTCACTGTCTGGGAGGAGAGA
GGGAAATGCCATGGGCCTACGAGCGAGTTTTCAAGTATGACATCACAGAGGTTGCAGATAGACTTA
AACTAAGTTATGGGGACACCTTTAACTTCCGACTAGAGATCACATCCTACGATGGATCGGTGGTAA
ACAAGAGCCTACCCAATCCTTTTCATCATCTACAGACCTGCCAATCATGACTACGATGTTCTTGTTA
TCCCAGTAGGAAGAAACCTTCACATCCCTCCCAAAGTTGTCGTCAAGAGAGGCACCCGCATCGAGT
TCCACCCAGTCGATGATTCACTTACGAGACCAGTTGTTGATCTTGGAAGCTACACTGCACTCTTCA
ACTGTGTGGTACCACCGTTTACATACCGCGGATTCGAAGTGAACCACGTCTATTCTGTCAAGCCTG
GTGACTACTATGTTACCGGACCAACGAGAGACCTTTGCCAGAATGCAGATGTCAGGATTCATATCC
ATGTTGAGGATGAGTAA

3'UTR

CGCAACAG

INTRON 3'UTR (SEQ ID NO:138)

GTGAGATAAGAAACCCTTCTAACAGTAATACGACACCACATTACAGCTTAAACATGATTGCCATCG
ATGTTTTTCATGTGTAGTATACGCTTTTTCAGTTCTACATAATTTTGTTTTTTCAAATCAAGTTTAGCA
AATGAATCTATCACTGGAAAATAGGGTAGGGTAGCCAAGTGGTTAAAGCGGTCACTGATCACGCCA
AAGACGAGTGTCTAACCTGCATGGGTACAAAAGTGAAGACCATTGCTGGTGTCTACCGCCGTAAT
ATTGTTTTTTAGTATTGCTAAAACCTTATACTCACCCTATGCGCTGTAAAAGTGGAATAATAATCATAT
TTCAACAAAAGCACAAAACCATTTTCATTTTCATGAAAGCCTCTTGTTACCTGAAAGACGCAAGAG
AACAATAGTTCCTAACATTATTTTCAGACATTGGAAATGTCCTGCACGTGTAAACCATATATCCTT
TGAAATTTTTTACGACTGCATCGTATACAATTTATGATATAAAATTTAAACCTTTATTTTCAG

3'UTR

GTTTCTTGGTCTCCACATATTCACACATCAGCACCAAACGGTTTTGAAGGACATTGGCGTTCTTCT
CTGGCAATGCATTTCAATACAACTTGAAGTGAAGTTCAGCATATCAGTGTGCTTCGAACGTGTTT
CGGAAGTACTCAAATGTGCTATGACTGAATTATTGTACATACATAACTTATTGATGTTCAATAAAT
AAATGTTGAAACG

Fig. 7a**Primary structure of the Hth2 protein****DOMAIN A (SEQ ID NO:156)**

GLPYWDWTQHLLTQLPDLVSDPLFVDPEGGKAHDNAWYRGNIKFENKKTARAVDDRLFEEKVGPGEN
RLFEGILDALEQDEFNCNFEIQFELAHNAIHVYLVGGRHTYSMSHLEYTSYDPLFFLHHSNTDRIFAI
WQRLQVLRGKDPNTADCAHNLIHEPMEPFRRDSNPLDLTRENSKPIDSFYAHGLGYQYDDLTLNGM
TPEELNSYLHERSGKEGVFASFRLSGFGGSANVVVYACRPAHDEMAVDQCDKAGDFFVLGGPTMP
WRFYRAFHFDVTDSDIDNIDKDRHGHYYVKAELFSVNGSALPNDLLPQPTISHRPARGHVDEAPAPS
SDAHLAVRKDINHLLTREEVYELRRAMERFQADTSVDGYQATVEYHGLPARCPFPEATNRFACCIHG
MATFPHW

DOMAIN B

HRLFVTQVEDALIRRGSPIGVPYWDWTQPM AHL PGLADNATYRDPISGDSRHNPFHDVEVAFENGR
TERHPDSRLFEQPLFGKHTRLFDSIVYAFEQEDFCDFEVQFEMTHNNIHAWIGGGGKYSMSSLHYT
AFDPISYLHHSNTDRLWAIWQALQIRRNKPYKAHCAWSEERQPLKPFASFSSPLNNNEKTYENSVPT
NVYDYEGLVGYTYDDLNF GGMDLGQLEEYIQRQRQDRFTFAGFFLSHIGTSANVEIIIDHGT LHTS
VGTF AVLGG EKEMKWGFDRLYKYEITDELRLQLNLRADDGFSISVKVTDVDGSELSSSELIPSAAIIF
ERSH

DOMAIN C

IDHQDPHQDTIIRKNVDNLTPEEINSLRRAMADLQSDKTAGGFQQIAAFHGEPKWCPSPDAEKKFS
CCVHGM AVFPWHRLTVQGENALRKHGCLGALPYWDWTRPLSHLPDLVSQQNYTDAISTVEARNP
WYSGHIDTVGVDTTTSVRQELYEAPGFGHYTGVAQVLLALEQDDFCDFEVQFEIAHNFIHALVGG
SEPYGMASLRYTTYDPIFYLHHSNTDRLWAIWQALQKYRGKPYNSANCAIASMRKPLQPFGLTDEI
NPDDETRQH AVPF SVFDYKNNFN EYDTLDFNGLSISQLDRELSRRKSHDRVFAGFLLHGIQQSAL
VKFFVCKSDDDCDHYAGEFYILGDEAEMPWG YDRLYKYEITEQLNALDLHIGDRFFIRYEA FDLHG
TSLGSNIFPKPSVIHDEGA

DOMAIN D

GHHQADEYDEVVTAASHIRKNLKDLSKGEVESLRSAFLQLQNDGVYENIAKFHGKPGLCDDNGRKV
ACCVHGMPTFPQWHRLYVLQVENALLERGS AVSVPYWDWTETFTELPSLIAEATYFNSRQQT FDPN
PFRGKISFENAVTTRDPQPELYVNRYYYQNVMLAFEQDNYCDFEIQFEMVHNVLHAWLGGRATYS
ISSLDYSAFDPVFFLHHANTDRLWAIWQELQRYRKKPYNEADCAINLMRKPLHPFDNSDLNHDPVT
FKYSKPTDGFYQNNFGYKYDNLEFNHFSIPRLEEIIIRIRQRQDRVFAGFLLHNIGTSATVEIFVC
VPTTSGEQNCENKAGTF AVLGG ETEMAFHFDRLYRFDISETLRDLGIQLDSHDFDLSIKIQGVNGS
YLDPHILPEPSLIFVPGSS

DOMAIN E

SFLRPDGHSSDDILVRKEVNSLT TRETASLIHALKSMQEDHSPDGFQAIASFHALPPLCPSPSATHR
YACCVHGMATFPQWHRLYTVQFQDALRRHGAAVGVPYWDWLRPQSHLP ELVTMETYHDIWSNRDFP
NPFYQANIEFEGENIT TEREVIADKLFVKGGHVFDNWFFKQAILALEQENYCDFEIQFEILHNGVH
TWVGGSRTHSIGHLHYASYDPLFY LHHSQTDRIWAIWQELQEQRGLSGDEAHCALEQMREPLK PFS
FGAPYNLNQLTQDFSRPEDTFDYRKFGYEYDNLEFLGMSVAELDQYIIIEHQENDRVFAGFLLSGFG
GSASVNFQVCRADSTCQDAGYFTVLGGSAEMAWAFDRLYKYDITETLEKMHLRYDDDF TISVSLTA
NNGTVLSSSLIPTPSVIFQRGH

Fig. 7bDOMAIN F

RDINTKSMSANRVRRELSDSL SARDPSSLKSALRDLQEDDGPNGYQALAAFHGLPAGCHDSQGNEIA
CCIHGMPTFPQWHRLYT LQLEMALRRHGSSVAIPYWDWTKPISELPSLFTSPEYYDPWHDVNNP
FSKGFVKFANTYTVRDPQEMLFQLCEHGESI LYEQTLLALEQTDYCDFEVQFEVLHNVIHYLVGGR
QTYALSSLHYASYDPFFFIHHSFVDKMVWVWQALQKRRKLPYKRADCAVNLMTKPMRPFDSMDNQN
PFTKMHAVPNTLYDYETLYSYDNLEIGGRNLDQLQAEIDRSRSHDRVFAGFLLRGIGTSADVRFW
ICRNENDCHRGGIIFILGGAKEMPWSFDRNFKFDITHVLEKAGISPEDVFDAAEPFYIKVEIHAVN
KTMIPSSVIPAPTI IYSPGE

DOMAIN G

GRAADSAHSANIAGSGVRKDVTTLT VSETENLRQALQGVIDDTGPNGYQAIASFHGSPPMCEMNGR
KVACCAHGMA SFPHWRLYVKQMEDALADHGSHIGIPYWDWTTAFTELPALVTDSENNPFHEGRID
HLGVTTSRSPRDM LFNDEPQGSSEFFYRQVLLALEQTDYCFEVQFELTHNAIHSWTGGRSPYGMS
TLEFTAYDPLFWLHHSNTDRIWAVWQALQKYRGLPYNEAHCEIQVLKQPLRPFND DINHNPITKTN
ARPIDSFDYERFNYQYDTLSFHGKSIPELNDLLEERKREERTFAAFLLRGIGCSADVVDICRPN
DCVFAGTFAVLGGELEMPWSFDRLFRYDITRVMNQLHLQYDSDFSFRVKLVATNGTELSSDLLKSP
TIEHEL

DOMAIN H

GAHRGPVEETE VTHQNTDGN AHFHRKEVDSL SLDEANNLKNALYKLQNDHSLTG YEAI SGYHGYPN
LCPEEGDDKYPCCVHGMAIFPHWRL LTIQLERALEHNGALLGV PYWDWTKDLSSLPAFFSDSSNN
NPYFKYHIAGVGHD TVREPTSLIYNQPQIHGYDYLYYLA LTTLEENNYCDFEVQYEILHNAVH SWL
GGSQKYSMTLEYS AFDPVFMILHSGLDRLWI IWQELQKIRRKPYNFAK CAYHMMEEPLAPFSYPS
INQDEFTRAN SKPSTVFD SHKFGYHYDNLNVRGHSIQELNTI INDLRNTDRIYAGFVLSGIGTSAS
VKIYLRTDDNDEEVGTFTVLGGEREMPWAYERVFKYDITEVADRLKLSYGDTFNFRL EITSYDGSV
VNKSLPNPFI IYR PANHDYDVLVI PVGRNLHI PPKVVVKRGTRIEFHPVDDSVTRPVVDLGSYTAL
FNCVVP PFTYRGFELNHVYSVKPGDYVVTGPTRDLCQNADVRIHIHVEDE

Fig 8a**Genomic sequence of the KLH1 gene**

DOMAIN 1B

GGCCTACCGTACTGGGACTGGACTGAACCCATGACACACATTCCGGGTCTGGCAGGAAACAAAACCT
TATGTGGATTCTCATGGTGCATCCACACAAATCCTTTTCATAGTTCAGTGATTGCATTTGAAGAA
AATGCTCCCCACACCAAAAGACAAATAGATCAAAGACTCTTTAAACCCGCTACCTTTGGACACCAC
ACAGACCTGTTCAACCAGATTTTGTATGCCTTTGAACAAGAAGATTACTGTGACTTTGAAGTCCAA
TTTGAGATTACCCATAACACGATTCACGCTTGGACAGGAGGAAGCGAACATTTCTCAATGTCTGTCC
CTACATTACACAGCTTTCGATCCTTTGTTTTACTTTTACCATTCTAACGTTGATCGTCTTTGGGCC
GTTTGGCAAGCCTTACAGATGAGACGGCATAAACCCCTACAGGGCCCCACTGCGCCATATCTCTGGAA
CATATGCATCTGAAACCATTTCGCTTTTTCATCTCCCCTTAACAATAACGAAAAGACTCATGCCAAT
GCCATGCCAAACAAGATCTACGACTATGAAAATGTCCTCCATTACACATACGAAGATTTAACATTT
GGAGGCATCTCTCTGGAAAACATAGAAAAGATGATCCACGAAAACCAGCAAGAAGACAGAATATAT
GCCGGTTTTCTCCTGGCTGGCATACTGTTTACGCAAATGTTGATATCTTCATTAATAACTACCGAT
TCCGTGCAACATAAGGCTGGAACATTTGCAGTGCTCGGTGGAAGCAAGGAAATGAAGTGGGGATTT
GATCGCGTTTTCAAGTTTGACATCACGCACGTTTTTGAAGATCTCGATCTCACTGCTGATGGCGAT
TTCGAAGTTACTGTTGACATCACTGAAGTCGATGGAATAAAGTTCATCCAGTCTTATTCCACAT
GCTTCTGTCATTTCGTGAGCATGCACGTGGTAAGCTGAATAGAG

INTRON 1B/1C (SEQ ID NO:139)

GTTTTGTAATAATTATGTAGAATTCTTTACCTCAGAATAAGATGAGGTCACATGGGTTTTGCAAAA
CTATTACGTTTCGAATTAATATTAATAATACCGGACCTCCACTGGTACATATTTATCTTTATAACG
ATAATAGCGATGATGATGATGATGATGATGATGATGATGATGATGATAATGATGATGCCGGTATTG
CACGTAATCCAGCCGACTTAGATGACACCCTAAGGGTGCAGAAAGTATAACAATTAGATTGCGTTT
GCATCTGTGTATGCGTGTGCTTTAACCAAAAAGTCAAATAAAAAGTGCAAACCTTAGTTTTATTCAT
TTGATAGAGCCTTTTACGATAAGAACAATGTAATAAATTAGAACATAACTGAAACCTCCGAAAGAA
GGCCTGTTTGTCAAGAGAGGTATCGACATGATTGACTTATAAACCTGTGCTTCTATATTTTGAAC
TGTCCACTTTCTTGTGTGTGTACTGTAATCACATCGCACTATGGCTGCAAGACGTGTACGAGTAC
ACTATATACTTACCTAATGACCAACCACAAGGCTGGCTTTGTTAATATTGTTATTTTACAGAAATA
AACACAGAATTCCAGCATTTGGCTGGTGTATTTAGCAAAACACCGATATGACACTCATGTTTTATT
ACATTTTTTTTCAG

DOMAIN 1C

TTAAATTTGACAAAGTGCCAAGGAGTCGTCTTATTCGAAAAAATGTAGACCGTTTGAGCCCCGAGG
AGATGAATGAACTTCGTAAAGCCCTAGCCTTACTGAAAGAGGACAAAAGTGCCGGTGGATTTTCAGC
AGCTTGGTGCATTCCATGGGGAGCCAAAATGGTGTCTTAGTCCCGAAGCATCTAAAAAATTTGCCT
GCTGTGTTACGGCATGTCTGTGTTCCCTCACTGGCATCGACTGTTGACGGTTCAGAGTGAAAATG
CTTTGAGACGACATGGCTACGATGGAGCTTTGCCGTACTGGGATTGGACCTCTCCTCTTAATCACC
TTCCCGAACTGGCAGATCATGAGAAGTACGTCGACCCTGAAGATGGGGTAGAGAAGCATAACCCTT
GGTTCGATGGTCATATAGATACAGTCGACAAAACAACAACAAGAAGTGTTTCAAGATAAAGTCTTCG
AACAGCCTGAGTTTGGTCAATTATACAAGCATTGCCAAACAAGTACTGCTAGCGTTGGAACAGGACA
ATTTCTGTGACTTTGAAATCCAATATGAGATTGCCATAACTACATCCATGCATTTGTAGGAGGCG
CTCAGCCTTATGGTATGGCATCGCTTCGCTACACTGCTTTTGATCCACTATTCTACTTGCATCACT
CTAATACAGATCGTATATGGGCAATATGGCAGGCTTTACAGAAGTACAGAGGAAAACCGTACAACG
TTGCTAACTGTGCTGTTACATCGATGAGAGAACCCTTTGCAACCATTGTCCTCTCTGCCAATATCA
ACACAGACCATGTAACCAAGGAGCATTCAAGTCCATTCAACGTTTTTTGATTACAAGACCAATTTCA
ATTATGAATATGACACTTTGGAATTTAACGGTCTCTCAATCTCTCAGTTGAATAAAAAGCTCGAAG

Fig. 8b

CGATAAAGAGCCAAGACAGGTTCTTTGCAGGCTTCCTGTTATCTGGTTTTCAAGAAATCATCTCTTG
TTAAATTCAATATTTGCACCGATAGCAGCAACTGTCACCCCGCTGGAGAGTTTTACCTTCTGGGTG
ATGAAAACGAGATGCCATGGGCATACGATAGAGTCTTCAAATATGACATAACCGAAAAACTCCACG
ATCTAAAGCTGCATGCAGAAGACCACTTCTACATTGACTATGAAGTATTTGACCTTAAACCAGCAA
GCCTGGGAAAAGATTTGTTCAAGCAGCCTTCAGTCATTCATGAACCAAGAATAG

INTRON 1C/1D (SEQ ID NO:140)

GTACTTGTTATATGTTTCGAATATTGCCGATACCTTCAATATATATACTTTATCAAAGTAATTGAT
TAATCTGAAGTAATTTTCCCTTCCAGTAGAGATTGAGTTGATACAACAAGAATTCGCCCTGTTGTA
TGTCACCTTTATTTTCATCAAACGATTGGAAGTGAGCTGTCCATGCCACAATGGGGTCTCTGTAAC
TTCTCGTATGGGGTATAGATTATATAGACGTGGCAGACCTTACGTATAACTAATATTTGTGTAATG
TCGTTTCAG

DOMAIN 1D

GTCACCATGAAGGCGAAGTATATCAAGCTGAAGTAACTTCTGCCAACCGTATTGAAAAAACATTG
AAAATCTGAGCCTTGGTGAACCTCGAAAGTCTGAGAGCTGCCTTCCTGGAAATTGAAAACGATGGAA
CTTACGAATCAATAGCTAAATTCATGGTAGCCCTGGTTTGTGCCAGTTAAATGGTAACCCCATCT
CTTGTTGTGTCCATGGCATGCCAACTTTCCCTCACTGGCACAGACTGTACGTGGTTGTCTGTTGAGA
ATGCCCTCCTGAAAAAAGGATCATCTGTAGCTGTTCCCTATTGGGACTGGACAAAACGAATCGAAC
ATTTACCTCACCTGATTTTCAGACGCCACTTACTACAATTCCAGGCAACATCACTATGAGACAAACC
CATTCCATCATGGCAAAATCACACACGAGAATGAAATCACTACTAGGGATCCCAAGGACAGCCTCT
TCCATTCAGACTACTTTTACGAGCAGGTCCTTTACGCCTTGGAGCAGGATAACTTCTGTGATTTG
AGATTGAGTTGGAGATATTACACAATGCATTGCATTCTTTACTTGGTGGCAAAGGTAAATATTCCA
TGTCAAACCTTGATTACGCTGCTTTTGATCCTGTGTTCTTCTTCATCACGCAACGACTGACAGAA
TCTGGGCAATCTGGCAAGACCTTCAGAGGTTCCGAAAACGGCCATACCGAGAAGCGAATTGCGCTA
TCCAATTGATGCACACGCCACTCCAGCCGTTTGATAAGAGCGACAACAATGACGAGGCAACGAAAA
CGCATGCCACTCCACATGATGGTTTTGAATATCAAAACAGCTTTGGTTATGCTTACGATAATCTGG
AACTGAATCACTACTCGATTCCCTCAGCTTGATCACATGCTGCAAGAAAGAAAAAGGCATGACAGAG
TATTCGCTGGCTTCCTCCTTCACAATATTGGAACATCTGCCGATGGCCATGTATTTGTATGTCTCC
CAACTGGGGAACACACGAAGGACTGCAGTCATGAGGCTGGTATGTTCTCCATCTTAGGCGGTCAA
CGGAGATGTCCTTTGTATTTGACAGACTTTACAACTTGACATAACTAAAGCCTTGAAAAAGAACG
GTGTGCACCTGCAAGGGGATTCGATCTGGAAATTGAGATTACGGCTGTGAATGGATCTCATCTAG
ACAGTCATGTATCCACTCTCCCACTATACTGTTTGAGGGCCGGAACAG

INTRON 1D/1E (SEQ ID NO:141)

GTAATATTTTGTCACTGTAACCAACAACCTGCAGTCTATTTTGCAATTACGATAATAACAATTTT
GAAATATATCTTTATTAAAGCAAAGGTTTCTAGAGACAAACAGCCGGCTCTAATTATTTTTTCGAA
CTTACGCTTGAGTAAAGATCTGCAAATGGCAACCCTACCTATACTATTAAAAATATAATGTTACAT
TCGTATCTGAATGTTTAATAAATCACTTCATATTCTGTTGCAG

DOMAIN 1E

ATTCTGCCCACACAGATGATGGACACACTGAACCAGTGATGATTTCGCAAAGATATCACACAATTGG
ACAAGCGTCAACAACCTGTCAGTGGTGAAGCCCTCGAGTCCATGAAAGCCGACCATTTCATCTGATG
GGTTCCAGGCAATCGCTTCCTTCCATGCTCTTCTCCTCTTTGTCCATCACCAGCTGCTTCAAAGA
GGTTTGCGTGCTGCGTCCATGGCATGGCAACGTTCCCAACAATGGCACCGTCTGTACACAGTCCAAT
TCCAAGATTCTCTCAGAAAACATGGTGCAGTCGTTGGACTTCCGTACTGGGACTGGACCCTACCTC

Fig. 8c

GTTCTGAATTACCAGAGCTCCTGACCGTCTCAACTATTTCATGACCCGGAGACAGGCAGAGATATAC
CAAATCCATTTATTGGTTCTAAAATAGAGTTTGAAGGAGAAAACGTACATACTAAAAGAGATATCA
ATAGGGATCGTCTCTTCCAGGGATCAACAAAAACACATCATAACTGGTTTATTGAGCAAGCACTGC
TTGCTCTTGAACAAACCACTACTGCGACTTCGAGGTTTCAGTTTGAAATTATGCATAATGGTGTTT
ATACCTGGGTTGGAGGCAAGGAGCCCTATGGAATTGGCCATCTGCATTATGCTTCCTATGATCCAC
TTTTCTACATCCATCACTCCCAAACCTGATCGTATTTGGGCTATATGGCAATCGTTGCAGCGTTTCA
GAGGACTTTCTGGATCTGAGGCTAACTGTGCTGTAAATCTCATGAAAACCTCCTCTGAAGCCTTTCA
GCTTTGGAGCACCATATAATCTTAATGATCACACGCATGATTTCTCAAAGCCTGAAGATACATTCG
ACTACCAAAAAGTTTGGATACATATATGACACTCTGGAATTTGCAGGGTGGTCAATTCGTGGCATTG
ACCATATTGTCCGTAACAGGCAGGAACATTCAGGGTCTTTGCCGGATTCTTGCTTGAAGGATTTG
GCACCTCTGCCACTGTTCGATTTCCAGGTCTGTTCGCACAGCGGGAGACTGTGAAGATGCAGGGTACT
TCACCGTGTTGGGAGGTGAAAAAGAAATGCCTTGGGCCTTTGATCGGCTTTACAAGTACGACATAA
CAGAAACCTTAGACAAGATGAACCTTCGACATGACGAAATCTTCAGATTGAAGTAACCATTACAT
CCTACGATGGAACGTACTCGATAGTGGCCTTATTCCACACCGTCAATCATCTATGATCCTGCTC
ATC

INTRON 1E/1F (SEQ ID NO:142)

GTAAGTATACACACATTATTTCTCTTCTGCTATATCAGATGAAGAGAACGTTGTATCACTAACCTA
GTCTTGTTTGATTTGTGGTTTCGTTTGCTTCCTGAACAGTAGGGTTGATTTAACTTCTCTGTTTCG
TCTGTACCAATGAAAGACTATGATGCTTGTGTGAAGATGCTTTGTTTCATGAGTCAGTCTGTTCTTG
TAATGCTTTGATCTTTGCCATCAACATTCTTGAAATTAATTATGGTTTCCCTTAAATACTTACATA
TTACATTTAAACGTCGCTGCTTGTCTGATTGCATATTCTTTCAAAAATAACTATATATTCAG

DOMAIN 1F-1 (1st part of domain f)

ATGATATTAGTTCGCACCACCTGTCGCTCAACAAGGTTTCGTTCATGATCTGAGTACACTGAGTGAGC
GAGATATTGGAAGCCTTAAATATGCTTTGAGCAGCTTGCAGGCAGATACCTCAGCAGATGGTTTTG
CTGCCATTGCATCCTTCCATGGTCTGCCTGCCAAATGTAATGACAGCCACAATAACGAG

INTRON 1F-1/1F-2 (SEQ ID NO:143)

GTAAATATACAGTGAAATCCGGATAAGTAAATCCAGATAAGAAAAAAACATTTTCTGTGGTCCC
GGCATGTTTCTTCTTCATCTATCATTATTTTGATACGGATAAGTAAAAATCGGCTGAGTAAACAT
CCGGGTAAGTAAATGATTTTCGAGGTCTCTTCATCGGATAAGTAAAGATACACAAGTGATCATTCC
AATAAACACTAACTGATGCAACACAATACCAGCGCACAGTGTTTTCACTACGTTTGTGTATTGT
AATTAACAATTAACACTTAAGTGTTTCCCAATGTGTCCGTGTGCAAACTGATTGGGACAAAGCTTG
CAACAAGCCCGCAATTCCATGTCGTTTATGTCTACGTTTGTATTCTGACTGCTTGGAGGGGTTT
GGAAAAAATAAAAAACGGGTAAATATTATAAAAAATTCACGGTGCCTTGAAATTTTAGGTGTCCG
GATTTCACTGTAGATGATTAATTTCTCACTTGTAACAAAAGGACCCAGTACCCTCATTTCGTGAC
GTACGTTATAAAATGTAATTATAAAAAGCCATTATCATGTTATACGTGATCTTGNCTTGCAATTA
TNCTACCGCTTTCTTGATTTTTTAAAGCAATTTCTCCCTCTATGAACTTATTAACATAGCACTCCT
GCAAAAGAAAACAGTCACTGCATGGATCCATATTGAATGTTGCTGCTTATTTCTCATTTTATTACT
CACAGATATTTCAAGAACATCGTACTCTTAACCAGGCTAAAGCAAAGAGGGTTACATTTTAGCCG
ACAAGTTCAGTACTGAGTGGAACACGTATATTAATGGAGATGACTCTGGTCATGATGATTAGG
ACAATTATCATGACGTTATCATTGATCATGACCATGTGAGTATAATAGATAGCTAACAAATAATGT
AATTACTAATTATGAAGCAATGGTGCATTTGCAG

DOMAIN 1F-2 (2nd part of domain f)

Fig. 8d

GTGGCATGCTGTATCCATGGAATGCCTACATTCCCCCACTGGCACAGACTCTACACCCTCCAATTT
GAGCAAGCTCTAAGAAGACATGGCTCTAGTGTAGCAGTACCCTACTGGGACTGGACAAAGCCAATA
CATAATATTCCACATCTGTTTCACAGACAAAGAATACTACGATGTCTGGAGAAATAAAGTAATGCCA
AATCCATTTGCCCCGAGGGTATGTCCCCTCACACGATACATACACGGTAAGAGACGTCCAAGAAGGC
CTGTTCCACCTGACATCAACGGGTGAACACTCAGCGCTTCTGAATCAAGCTCTTTTGGCGCTGGAA
CAGCAGCACTACTGCGATTTTGCAGTCCAGTTTGAAGTCATGCACAACACAATCCATTACCTAGTG
GGAGGACCTCAAGTCTATTCTTTGTCATCCCTTCATTATGCTTCATATGATCCGATCTTCTTCATA
CACCCTCCTTTGTAGACAAGGTTTGGGCTGTCTGGCAGGCTCTTCAAGAAAAGAGAGGCCTTCCA
TCAGACCGTGCTGACTGCGCTGTTAGTCTGATGACTCAGAACATGAGGCCTTTCCATTACGAAATT
AACCATAACCAGTTTACCAAGAAACATGCAGTTCCAAATGATGTTTTCAAGTACGAACTCCTGGGT
TACAGATACGACAATCTGGAAATCGGTGGCATGAATTTGCATGAAATTGAAAAGGAAATCAAAGAC
AAACAGCACCATGTGAGAGTGTTTGCAGGGTTCTCTCCTTCACGGAATTAGAACCTCAGCTGATGTC
CAATTCCAGATTTGTAAACATCAGAAGATTGTCACCATGGAGGCCAAATCTTCGTTCTTGGGGGG
ACTAAAGAGATGGCCTGGGCTTATAACCGTTTATTCAAGTACGATATTACCCATGCTCTTCATGAC
GCACACATCACTCCAGAAGACGTATTCCATCCCTCTGAACCATTCTTCATCAAGGTGTCAGTGACA
GCCGTCAACGGAACAGTTCTTCCGGCTTCAATCCTGCATGCACCAACCATTATCTATGAACCTGGT
CTCGGTG

INTRON 1F-2/1G-1 (SEQ ID NO:144)

GTCTCGGTGAGTTATTAAAAGAAACAAAATATTTACCATTACCATTGTTAACTACAAAAATGAGTG
AGATATCTTATATCACTGGTACACTACTGATATTTTATGCAATGAAATTACTATTTTTTCCAGGTAC
GCTTCAACCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCATCATGCTTTTCTGT
AAAACATAAAACACCAATTAACAATGTTCTTAGTGTGTTTGTGACTCCCTTCCACTGCAACGCCT
ACATAATCAAAGTGTTCTGTTTTTTTCCAACTTTCCAGTTAGTGTGTTGAAGACTAAAAAGTTAAATA
AGCATTACATAACTTCTAAGAGCAACTGGGACCATGCAGTTACGTATTGATATTTCTGTGAGAGT
GAAGCAAAACACTGTTTTTTCAAGCTTAGGTTTATCAATCAAATGTCCAATAGTTCATGTTATCGA
AAAGGCAGCGAAGGATAAGAGGCTCCGAGACATCTTGTCTATTCTCGTGTTTCATATGATATCAACT
GAGGAGCTTCCATTACATTTTTTGACCTTATCATTTAAAGACATACATGGAACATTTTCATTTTACA
GTTAAAGTGAACCACTTCAGGTTCAACTTCAACTTCGAATTCAACTTCTGTTGTGTGTTTTATGAG
CCGACTGAAATAGAGTGCCTTACTTTCACTTCTAGTTTCGTTCTGTCTCGTCATCGTTGTTTCTTT
CAGTGTGCATAGTACACGCCTAGTATAGAACACACGAACCTTGTCCTTACTTAATAGATTCTGAAAC
TATTATGTGGAAAGTTGGCAGGCTATAGTAACATCCTGGCAAAATTATCATGTATCCTCTTGTTTG
TCATAATTAG

DOMAIN 1G-1 (1st part of domain g)

ACCATCACGAAGATCATCATTTCTTCTTCTATGGCTGGACATGGTGTCAGAAAGGAAATCAACACAC
TTACCACTGCAGAGGTGGACAATCTCAAAGATGCCATGAGAGCCGTCATGGCAGACCACGGTCCAA
ATGGATACCAGGCTATAGCAGCGTTCCATGGAAACCCACCAATGTGCCCTATGCCAGATGGAAAGA
ATTACTCGTGTTGTACACATG

INTRON G1-1/1G-2 (SEQ ID NO:145)

GTATGTATTTCCCACTGGTGGTCGCTGACTGCCAACACATACTTGTAATTTATTCATGAAAGTATA
ATAGTTTGTGTTGAAAGTATATTTATAACCATCTTGACACAAGCGTCACGAATTTTACCACAAAGCT
TCAAAACGCCCAAAACATTCTAATAGCGATATATTTGTTAAAAGACCAAAATATAGCCTTACAACA
ATAGATTATTTTAATAAGACCAGTCAGTGCATGCAATCGATTGGAAACTTTGAAATAAAATATTC
TATGTACTAACTGCCAATCTCATAATACTTGCCCTTGATGTGCTTCTTTTTTCACATTCGCGTCGAG
CTTCAACTCCAATGCATAAGCTTAAAAATAATCATAAACACAAACAAATAGCCACAGAGGCGACGA
TCCCTCCAGGCCAGGCTTTATTTGTCTCTTATAGAATATATCGCTATTAGAATGTTTTTGACGTTT

Fig. 8e

TGAAGCTTTGTGGGTGAAAATTCGTGATGTTTATGCGTGGTATTTATGTAAGATGAAAATAAATAT
ATCTTTTCAAACAAGATTTTAGTATTTTGAAGACTTCTATGAATAAATTACACTTATGTGTTAGGT
TATTGGTCACTGAGCGCTTGTGGTATTTTCCCTTCTTCAATTTGTTTGTCTTTGTTCAATTTCGA
ATAGTTATCCTACTGTGGATAGTCTATATGAGAATCGTTGAAAGAATAATACAATTCTAATGGATT
GCAACTTCTTTAACTTTTATTTGCAACTGCCACGTTTCGGTATACGTTCTTATGCCGTCATCAAGC
ATACGAGTGTACATGTATGCCAAAACGCTGCAATAAAAATTAAAGAAGTTGCAATCCATAAGAAT
TTCAATGTTCTTTTCATCATCACATCAACTTCTAAAAATGCCTATAAAAACAATCAACAAACGTACAA
TAGTACATTACCGGATCTCGCAGCATGACCACGTCGATATCTAAACAATATCACTATCCATTAATA
GGATCAAGAGTAGGTACAGACATGTTTCAAGTTATAAATACTCTTCAAAAAAGTAGGGGAACTTGGA
TTTCAAGGTCAATAACAACTAATGATAATAACAATTGGTCCCAAATAATAACAATTGGTCCCAA
CTAATTGTATCTTTACAAAGAAGAAATTGAGTGAACAATTCACCCGGTATTTTATTACCTAAACCG
TTTCTCTTGCTGTTATGGTGCGTGAAAGAAGAAATGGGTAAGAAACGGAAATTGACATTTTTCGCT
CAGTGGTGCGTAATGCCCCCATTGTTGGCCAAACACTGATTGATTGCTGAGGCATCGTGCATACG
CGTCTACCTATGGTAATTTGATGCAGTCTGTCCATTCTTCCACCAACGCCTGGACAAGTTCATCT
AGCGTGGCTGGTGGCCTTTACGTTGACGCACACGTCGGCCCAAGATGTCCAGACATTTTCAATG
GCCAGGGCTCATTGCTGGTCAGGGCATCCTATGGATATTGTGCCGTTGAAGGTGGTTATGTTGTTT
ACATTGAAATTCCAAGTTCTCCTACTCTTTTAAAGAGGAGGTTACAAAGTACGTTCTTTTCATGTT
GGTGAAGAGAATATCAAGGTCTTCTAAGGGATTGTGTCTTATAATATTTGATTTTAAAGAAGTTGA
TATTATCTGCATCCTTCCCAAGAAATTGCAATGTTTACACACTATTGCGTTTGATAATGTTTTTG
GGGAAATAAAGTGTCCAGGACTGCTAAATAGTAATTATTGCTACTTTTAG

DOMAIN 1G-2 (2nd part of domain g)

GCATGGCTACTTTCCCCCACTGGCACAGACTGTACACAAAACAGATGGAAGATGCCTTGACCGCCC
ATGGTGCCAGAGTCGGCCTTCTTACTGGGACGGGACAACCTGCCTTTACAGCTTTGCCAACTTTTG
TCACAGATGAAGAGGACAATCCTTTCCATCAT

INTRON 1G-2/1G-3 (SEQ ID NO:146)

GTGAGTTCACGTAAGCCTACGAGATCAACATTACTCCTTAACAGCCACGGCATCATGTACCGATAT
ATCACAACAAAAGTATTCAAAGCTTTAAACACGATATGTATGGTTCAGAATGACATCATTAAC
AAGGACATGAGTCTGAAATAAACATGACTTGACACCGTTGTGGTCACAGTTTTGTTTCTCATTGGT
GAACCTGTGAAACAACCTTTCAAACCAAAAGATGCCTATTAATATTGTTAATTCCCATGAATTAGG
AGATACACACATTCTACTGTCATTT.....AATAACCGCTTC
CAGCATGAAAACACAATATGATTATCTCAATTCTACCATTACTAATTATAATTTTGACTGGCATT
TTTGACGACGCGTAAACATCGCTGCTTTACAGACTGCACTGCGGTAACGTGACGTTTTTCATGAC
GTCACTACATTCTATTCAAACATTTCCACAGAAGAGCGAGACCACGGCCGTGATGGGTTCTGGGC
AGATGATTACCCAAGTATATATTTATAATAACTTGACTGCTTGCCCTGAATAATGTTGACACATGAC
AACGAATTTGTGATAGCGTAAGAAGCGTGAATACTGTGAATAGTGTGAGGGGTGTTTGCTGAGAGT
TAACCACCGTTAATTGCAAAAATTCCTGAATACCTTGCAATTTGCAGTCAAGAAGAATTGCATTCTTA
CTCCTGTGAATGGACTCATTGTTATTTAGCAGCGGTTATTGAGGTTTTGATCACCTCTAAATAGAC
AATCAGGATGCGGCAACCGGAAAATTATAGCAGAATCTGTAATTCAAGATGGGCTTGCCCTGTGAA
AATATGCTGCGAGTTCAGTAACACTTTTCCCTTTTCGATCATGGCCTGTTTTGCTCTGAATCTGGTC
TTTCAGAGGATCCCTGCTTTTTTAAACTAAAGTCTTCCCAACTCACTTATATTTATGTTTTTTAA
TTATTTATAGTTTTTAATATGAACAACAAATCATATTTATTTACACATTATATTTTTTCAG

DOMAIN 1G-3 (3rd part of domain g)

GGTCACATAGACTATTTGGGAGTGGATACAACCTCGGTGCCCCGAGACAAGTTGTTCAATGATCCA
GAGCGAGGATCAGAATCGTTCTTCTACAGGCAGGTTCTCTTGGCTTTGGAGCAGACAGAT

Fig. 9a**Primary structure of the KLH1 protein****DOMAIN B**

GLPYWDWTEPMTHIPGLAGNKTYVDSHGASHTNPFHSSVIAFEENAPHTKRQIDQRLFKPATFGHH
TDLFNQILYAFEQEDYCDFEVQFEITHNTIHAWTGGSEHFSMSSLHYTAFDPLFYFHHSNVDRDLWA
VWQALQMRRHKPYRAHCAISLEHMLKPPAFSSPLNNNEKTHANAMPNKIYDYENVLHYTYEDLTF
GGISLENI EKMIHENQQEDRIYAGFLLAGIRTSANVDIFIKT TDSVQHKAGTFAVLGGSKEMKWGF
DRVFKFDITHVLKDLDLTADGDFEVTVDITEVDGTKLASSLI PHASVIREHARGKLN R

DOMAIN C

VKFDKVP RSRLIRKNVDRLSPEEMNELRKALALLKEDKSAGGFQQLGAFHGEPKWCPSP EASKKFA
CCVHGMSVFPWHRLTLTVQSENALRRHGYDGALPYWDWTSPLNHLPELADHEKYVDPEDGVEKHNP
WFDGHIDTVDKTTTRSVQNKLF EQPEFGHYTSIAKQVLLALEQDNFCDFEIQYEIAHNYIHALVGG
AQPYGMASLRYTAFDPLFY LHHSNTDRIWAIWQALQKYRGKPYNVANCAVTSMREPLQPFGLSANI
NTDHVTKEHSVPFNVDYKTNFNYEYDTLEFNGLSISQLNKKLEAIKSQDRFFAGFLLSGFKKSSL
VKFNICTDSSNCHPAGEFYLLGDENEMPWAYDRVFKYDITEKLHDLKLHAEDHFIYIDYEVFDLKPA
SLGKDLFKQPSVIEPRI

DOMAIN D

GHHEGEVYQAEVTSANRIRKNIENLSLGELESRAAFLEIENDGTYESIAKFHGSPGLCQLNGNPI
SCCVHGMPTFPWHRLYVVVVENALLKKGSSVAVPYWDWTKRIEHLPHLISDATYYNSRQH HYE TN
PFHHGKI THENEITTRDPKDSL FHSDFYEQVLYALEQDNFCDFEIQLEILHNALHSL LGGKGKYS
MSNLDYAAFDPVFFLHHATTDRIWAIWQDLQRFRKRPYREANCAIQLMHTPLQPFDKSDNNDEATK
THATPHDGFYQNSFGYAYDNLELNHYSIQLDHMLQERKRHDRV FAGFLLHNIGTSADGHV FVCL
PTGEHTKDCSHEAGMFSILGGQTEMSFVFDRLYKLDITKALKKNGVHLQGD FDL EIEITAVNGSHL
DSHVIHSP TILFEAG

DOMAIN E

TDSAHTDDGHTEPVMIRKDI TQLDKRQQLSLVKALESMKADHSSDGFQAIASFHALPPLCPSPAAS
KRFACCVHGMATFPQWHRLYTVQFQDSL RKHGAVVGLPYWDWTLPRSELPELLTVSTIHD PETGRD
IPNPFIGSKIEFEGENVHTKRDINRDLRFQGSTKTHHNWFIEQALLALEQTNYCDFEVQFEIMHNG
VHTWVGKPEPYGIGHLHYASYDPLFYIHHSQTDRIWAIWQSLQRFRGLSGSEANCAVNLMKTPLKP
FSFGAPYNLNDHTHDFSKPEDTFDYQKFGYIYDTLEFAGWSIRGIDHIVRNRQEHSRVFAGFLLEG
FGTSATVDFQVCRTAGDCEDAGYFTVLGGEKEMPWAFDRLYKYDITETLDKMNL RHDEIFQIEVTI
TSYDGTVLDSGLIPTPSIIYDPAH

DOMAIN F

HDISSHLSL NKVRHDLSTL SERDIGSLKYALSSLQADTSADGFAAIASFHGLPAKCND SHNNEVA
CCIHGMP TFPWHRLYTLQFEQALRRHGSSVAVPYWDWTKPIHNI PHLFTDKEYYDVWRNKVMPNP
FARGYVPSHDTYTVRDVQEGFLH LTSTGEHSALLNQALLALEQHDYCDFAVQFEVMHNTIHYLVGG
PQVYSSL SHYASYDPIFFIHHSFVDKVWAVWQALQEKRLPSDRADCAVSLMTQNMRPFHYEINH
NQFTKKHAVPNDVFKYELLGYRYDNLEIGGMNLHEIEKEIKDKQHHVRVFAGFLLHGIRTSADVQF
QICKTSEDCHHGGQIFVLGGTKEMAWAYNRLFKYDITHALHDAHITPEDVFHPSEPF FIKVSVTAV
NGTVLPASILHAPTIIYEPGLG

Fig. 9b

DOMAIN G

DHHEDHHSSSMAGHGVRKEINTLTAEVDNLKDAMRAVMADHGPNQYQAIAAFHGNPPMCPMPDGK
NYSCCTHGMATFPHWHRLYTKQMEDALTAHGARVGLPYWDGTTAFTALPTFVTDEEDNPFHHGHID
YLGVDTTTRSPRDKLFNDPERGSESFFYRQVLLALEQTD

Fig. 10a**Genomic sequence of the KLH2 gene****DOMAIN 2B**

GGCCTGCCCTACTGGGATTGGACCATGCCAATGAGTCATTTGCCAGAACTGGCTACAAGTGAGACC
TACCTCGATCCAGTTACTGGGGAACTAAAAACAACCCCTTTCCATCACGCCCAAGTGGCGTTTGAA
AATGGTGTAAACAAGCAGGAATCCTGATGCCAACTTTTTATGAAACCAACTTACGGAGACCACACT
TACCTCTTCGACAGCATGATCTACGCATTTGAGCAGGAAGACTTCTGCGACTTTGAAGTCCAATAT
GAGCTCACGCATAATGCAATACATGCATGGGTTGGAGGCAGTGAAAAGTATTCAATGTCTTCTCTT
CACTACACTGCTTTTGATCCTATATTTTACCTCCATCACTCAAATGTTGATCGTCTCTGGGCCATT
TGGCAAGCTCTTCAAATCAGGAGAGGCAAGTCTTACAAGGCCCACTGCGCCTCGTCTCAAGAAAGA
GAACCATTAAGCCTTTTGCATTTCAGTTCCCCACTGAACAACAACGAGAAAACGTACCACAACCTCT
GTCCCCACTAACGTTTATGACTATGTGGGAGTTTTCGACTATCGATATGATGACCTTCAGTTTGGC
GGTATGACCATGTCAGAACTTGAGGAATATATTCAACAAGCAGACACAACATGATAGAACCTTTGCA
GGATTCTTCCTTTCATATATTGGAACATCAGCAAGCGTAGATATCTTCATCAATCGAGAAGGTCAT
GATAAATACAAAGTGGGAAGTTTTGTAGTACTTGGTGGATCCAAAGAAATGAAATGGGGCTTTGAT
AGAATGTACAAGTATGAGATCACTGAGGCTCTGAAGACGCTGAATGTTGCAGTGGATGATGGGTTT
AGCATTACTGTTGAGATCACCGATGTTGATGGATCTCCCCCATCTGCAGATCTCATTCCACCTCCT
GCTATAATCTTTGACGTGGTCAGAG

INTRON 2B/2C (SEQ ID NO:147)

GTATTTAAAAAAGTAATAAAACCATATTTTCGAATGCGCTTTATGAAATATCGTGTGACTGGTTCT
TTAGTTTACATGGAGTGTAAACAACATGCTCCATCAGTTGACATATACTGCTCACACAAAGTAAGGG
ATATTTGATAATGATAACAAATATAATCAAAGCGGTTATACTATCAAGACTTATTCACATAATTAC
AGGTGAAGGGAGGTGTGATCGTGTTCACTGATCAGGTTGAGGCCAGAGAAGTCCCAGTTTGAGTCT
TGCAGAAGATGATGTTTAGGCATGGGGTTCGAATCACCAAAATCACATGACTTCAATAACGGGTGG
ACCACCTCGAGCGACGATGCAAGCAGTAGAGCGTCTACGCATGCTCCTGATAAGGCGACCAATCTG
TTCCTGGGGAATCAGTCGCCACTCCTCTTGTAGTGCCACGCTCATTTCTGCTACGGTCTGGGTAC
CTGCTATCGGGTCTTGATCCGTATCCCAAGGATGTCCACACATGTTCAAGGTGAGAGGTTCGGGGA
ACATCGCTGGCCACGGTAAGGTCTGAATTTGATGCCGTTGAAAGTGAGCTCTGACAACCTGAGCAT
GGTGAGCTCTGACGTTGTCTGCTGAAAGATGAATCCAGCTCCATGACAGCGAGCAAAGGGCAGGA
CGTGTTGGTCAATGCAGTTGTCTCTGCAGTACACACCTGTCACTCGCCACTCACAAGCGTGTAGAT
CTGTACGACCAGTCATGGAGATCCCAGCCCACATCATAACGGACCCCTATCCATACCGATCATGAG
CCACCATAGCAGCGTCTTGATGACGTTCTCCCTGTGCGCTCGACATCCTCACACGGCCAAAAGGAA
CGTGGAAGTCTGACTGAACATGACATTAGCCAACCTGGCACTTGTCCACCGCTGATGTTGGCGAGA
CCATTCCAGTCGAGCTCTTCGGTGTCTGGCTTTCATCGATAACACGACGTAAGGTCTGCGGGCGTG
CAAGACGGCTCTATGCAGGCGATTTTCGGATTGTCTGGGTGCTAACTCTGATCCAGGTGCCTGCTG
AAGTTGATGCTGGATCTGTGTGGCATTGAGATGGCGATTCTTAGGACTGTGGAGATGATGAATCG
ATCTTGACTTATGGTGGTGACATTAGGACGTCGGGTTCTGTCTCTATCCTGCACTCTTCCAGTTGT
TCGGTGACGCTCTGGTACCCGGCTGATTACTGACTGAGAATATCCATCTGCCGTGCGACATGAGCC
TGTGTTGGCCCAGCCTGAAGCATTGCAATCGCCAGAGACGCTCTTCAAAGTCAATTCGACGCATGG
TTTTCTGTTTCAAAATGACAGCGTAAACAGTTTTTGGTGCTTTTATGCTTCCCAAGAGCATGAAA
AACACGTTCTATGGGTCGTGCACACCTTACATGACAAGTGTGAAAAGTGACTTGCACCCCCCTTGTG
TGTTTCGGATGCACACTCTGTTTACGTACTGATGCGATTTGGCGTCTAAACATGTTTTGGCGTCTAA
ACATGTTTTCTGTCATGATTCATATACTATTTTGTTCATATTCCTGGCATCAAACCAAACCTACAGTG
AAATATATTTCAATATCCCCTACTTTGTGTGAGTAGTATAGATCACTGCAGACAACATATAGACAA
TGCAGTTACACCGTCAACAATCCCAGTCATTAATTATGATGACACTTCCACACATAGTGTGAGTGA
TTGTAATTCAACTGTACACACTTTTCCCGTGAACATTCAGGATCTATATGACTAAATATATAACAT
TAGTATACGTGCAGTTTTGTATCGCTACGACATTGTTGTAACCTTTGTTTAAATCATTTAACAG

Fig. 10b

DOMAIN 2C

CTGATGCCAAAGACTTTGGCCATAGCAGAAAAATCAGGAAAGCCGTTGATTCTCTGACAGTCGAAG
AACAAACTTCGTTGAGGCGAGCTATGGCAGATCTACAGGACGACAAAACATCAGGGGGTTTCCAGC
AGATTGCAGCATTTCCACGGAGAACC AAAATGGTGTCCAAGCCCCGAAGCGGAGAAAAAATTTGCAT
GCTGTGTTTCATGGAATGGCTGTTTTCCCTCACTGGCACAGATTGCTGACAGTTCAAGGAGAAAATG
CTCTGAGGAAACATGGATTTACTGGTGGATTGCCCTATTGGGACTGGACTCGGCCAATGAGCGCCC
TTCCACATTTTGTGCTGATCCTACTTACAATGATTCTGTTTCCAGCCTCGAAGAAGATAACCCAT
GGTATCATGGTTCACATAGATTCTGTTGGGCATGATACTACAAGAGCTGTGCGTGATGATCTTTATC
AATCTCCTGGTTTTCGGTCACTACACAGATATTGCAAAACAAGTCCTTCTGGCCTTTGAGCAGGACG
ATTTCTGTGATTTTGAGGTACAATTTGAAATTGCCATAATTTCATACATGCTCTGGTTGGTGGTA
ACGAACCATAACAGTATGTCATCTTTGAGGTATACTACATACGATCCAATCTTCTTCTTGACCGCT
CCAATACAGACCGACTTTGGGCCATTTGGCAAGCTTTGCAAAAATACCGGGGGAAACCATAACA
CTGCAAACTGTGCCATTGCATCCATGAGAAAACCACTTCAGCCATTTGGTCTTGATAGTGTCTATA
ATCCAGATGACGAACTCGTGAACATTCGGTTCCTTTCCGAGTCTTCGACTACAAGAACAACCTTCG
ACTATGAGTATGAGAGCCTGGCATTTAATGGTCTGTCTATTGCCCAACTGGACCGAGAGTTGCAGA
GAAGAAAGTCACATGACAGAGTCTTTGCAGGATTCCTTCTTCATGAAATTGGACAGTCTGCACTCG
TGAAATTCTACGTTTGCAACACAATGTATCTGACTGTGACCATTATGCTGGAGAATTCTACATTT
TGGGAGATGAAGCTGAGATGCCTTGGAGGTATGACCGTGTGTACAAGTACGAGATAACACAGCAGC
TGCACGATTTAGATCTACATGTTGGAGATAATTTCTTCCTTAAATATGAAGCCTTTGATCTGAATG
GCGGAAGTCTTGGTGAAGTATCTTTTCTCAGCCTTCGGTGATTTTCGAGCCAGCTGCAG

INTRON 2C/2D (SEQ ID NO:148)

GTATGTTTTAAATGTCACCTTATCCGTGATCTGTAATGAAGTTAGCAATTCACCTTATCAACTGTTT
GGCTGTACTGTTTCAGTGCGAGTTTTACTTAGGTTGGATTAATTAAAATATTCAAGCTCATAAATG
TTTTGATTCAACTTTTGTATTTATTTCAAACAG

DOMAIN 2D

GTTACACCAGGCTGATGAATATCGTGAGGCAGTAACAAGCGCTAGCCACATAAGAAAAAATATCC
GGGACCTCTCAGAGGGAGAAATTGAGAGCATCAGATCTGCTTTCCTCCAAATTCAAAAAGAGGGTA
TATATGAAAACATTGCAAAGTTCCATGGAAAACCAAGACTTTGTGAACATGATGGACATCCTGTTG
CTTGTTGTGTCCATGGCATGCCCACCTTTCCCCACTGGCACAGACTGTACGTTCTTCAGGTGGAGA
ATGCGCTCTTAGAACGAGGGTCTGCAGTTGCTGTTCTTACTGGGACTGGACCGAGAAAGCTGACT
CTCTGCCATCATTAATCAATGATGCAACTTATTTCAATTCACGATCCCAGACCTTTGATCCTAATC
CTTCTTCAGGGGACATATTGCCTTCGAGAATGCTGTGACGTCCAGAGATCCTCAGCCAGAACTAT
GGGACAATAAGGACTTCTACGAGAATGTATGCTGGCTCTTGAGCAAGACAACTTCTGTGACTTTG
AGATTACGCTTGAGCTGATACACAACGCCCTTCATTCTAGACTTGAGAGGAAGGGCTAAATACTCCC
TTTCGTCTCTTGATTATACCGCATTTGATCCTGTATTTTTCTTCCATTGCAACGTTGACAGAA
TCTGGGCCATCTGGCAGGACTTGCAGAGATATAGAAAGAAACCATACAATGAGGCTGACTGCGCAG
TCAACGAGATGCGTAAACCTCTTCAACCATTTAATAACCCAGAACTTAACAGTGATTCCATGACGC
TTAAACACAACCTCCCACAAGACAGTTTTTGATTATCAAAACCGCTTCAGGTACCAATATGATAACC
TTCAATTTAACCCTTCAGCATACAAAAGCTAGACCAAATTTAGGCTAGAAAACAACACGACA
GAGTTTTTGTGCTGGCTTTATTTCTTCACAACATTGGGACATCTGCTGTTGTAGATATTTATATTTGCG
TTGAACAAGGAGGAGAAACAAAACCTGCAAGACAAAGGCGGGTTCCTTCACGATTCTGGGGGGAGAAA
CAGAAATGCCATTCCACTTTGACCGCTTGTAACAATTTGACATAACGTCTGCTCTGCATAAACTTG
GTGTTCCCTTGGACGGACATGGATTTCGACATCAAAGTTGACGTGAGAGCTGTCAATGGATCGCATC
TTGATCAACACATCCTCAACGAACCGAGTCTGCTTTTTGTTTCTGGTGAACGTAAGAATATATATT
ATG

Fig. 10c

INTRON 2D/2E (SEQ ID NO:149)

GTTATAAAGCAGTATATTCTCTTCAAAAAAGTAGGGGAACCTTGGAATTTCAAGGTAAATAACATAA
CTACCTTCAACGGCACAATATCCATATGATGCCCTGGCCAGCAATGAGGCCTGATCTTTTCCCAT
TAAAAATGTCTGGAACATCTTGGGCAAACGTGTGCGTCAACGTAAAACGCCACCAGTCACGCTAGA
TGAACCTGTCCAGGCGTTGGTGGAAGAATGGGACAGACTGCATCAATTACCATAAGTAGACTCATT
TGCAGCGAATCAGTCAGTGTTTGACCAATAACGGGGGCATTACGCACTACTGACGCAAAACAATGT
CAATTTCCGTTTCTTACCCATTCTTTCTTTACGGACCATAACAGCAAGAGAACTGNTTAGGTAA
TGAAATACCGGTGAATTATTGTTAACTGGATTCTTCTTTGTAAAGATAACAATTAGTTTGGGACCA
ATTATTATTATCATTAGTTTGTATTGACCTTGAAATTCGAAGTTCCTCTACATTTTTTAAGGAGT
TTATTTGATTGACAATGAAATGTAAGAAAAGAGCAAATCGTAAATACGTTAAAAATTATTCCTTA
AACATCAGTCTCTAACTTCAGTTTAAATTGCCAGTAACACGTGTTATATGATGTTTCCGTTTCTCT
TTGTTTTTTTAGCATTCAACTTATTTGATATAACGTTTTACTGTTTTAGATTACATCAAACCTGCAG

DOMAIN 2E

ATGGGCTTTTCAACAATAATCTTGTGCGAAAAGAAGTAAGCTCTCTTACAACACTGGAGAAACATT
TTTTGAGGAAAGCTCTCAAGAACATGCAAGCAGATGATTCTCCAGACGGATATCAAGCTATTGCTT
CTTTCCACGCTTTGCCTCCTCTTTGTCCAAGTCCATCTGCTGCACATAGACACGCTTGTTGCCTCC
ATGGTATGGCTACCTTCCCTCAGTGGCACAGACTCTACACAGTTCAGTTCGAAGATTCTTTGAAAC
GACATGGTTCTATTGTCTGGACTTCCATATTGGGATTGGCTGAAACCGCAGTCTGCACTCCCTGATT
TGGTGACACAGGAGACATACGAGCACCTGTTTTACACAAAACCTTCCCAAATCCGTTCTCTCAAGG
CAAATATAGAATTTGAGGGAGAGGGAGTAACAACAGAGAGGGATGTTGATGCTGAACACCTCTTTG
CAAAGGAAATCTGGTTTACAACAACCTGGTTTTGCAATCAGGCACTATATGCACTAGAACAAGAAA
ATTACTGTGACTTTGAAATACAGTTCGAAATTTTGCAATAATGGAATTCATTCATGGGTTGGAGGAT
CAAAGACCCATTCAATAGGTCATCTTCATTACGCATCATAACGATCCACTGTTCTATATCCACCATT
CGCAGACAGATCGCATTGTTGGGCTATCTGGCAAGCTCTCCAGGAGCACAGAGGTCTTTCAGGGAAGG
AAGCACACTGCGCCCTGGAGCAAATGAAAGACCCTCTCAAACCTTTCAGCTTTGGAAGTCCCTATA
ATTTGAACAAACGCACTCAAGAGTTCTCCAAGCCTGAAGACACATTTGATTATCACCGATTCTGGGT
ATGAGTATGATTCCCTCGAATTTGTTGGCATGTCTGTTTCAAGTTTACATAACTATATAAAACAAC
AACAGGAAGCTGATAGAGTCTTCGCAGGATTCTTCTTAAAGGATTGGAACAATCAGCATCCGTAT
CGTTTGATATCTGCAGACCAGACCAGAGTTGCCAAGAAGCTGGATACTTCTCAGTTCTCGGTGGAA
GTTTCAGAAATGCCGTGGCAGTTTGACAGGCTTTACAAGTACGACATTACAAAAACGTTGAAAGACA
TGAAACTGCGATACGATGACACATTTACCATCAAGGTTACATAAAGGATATAGCTGGAGCTGAGT
TGGACAGCGATCTGATTCCAACCTCTTCTGTTCTCCTGAAGAAGGAAAGC

INTRON 2E/2F (SEQ ID NO:150)

GTATGTATCTCATGTTTCTCAAATAATTTGATTTTCAATGCCCTTACTATAAAGCACAGTTATTGT
TCAGTGCCAGTAACCGTTTATTTACGTAAATGTTACAGGCTATTATAATCAAAAATACATTACCGA
TATTGTTTACCACACAATTATATCATTGTCAAAATCTACCCCCATTACCTGCGTTTTGAATTTGTA
ACCTTCTGACAAAAATGAATTAGCAAGAGCTCTGATGAAGAACATAATGAACAACACCTATCTTTC
TTCTTTCAATGACGGTTTAACAATACAATGCACAATGTAAAAAATATATATATATATAATTTT
ATATCTACAGTTAATGCAAATGACTCCACTAATTCAGGGAAACACATTTTCAG

DOMAIN 2F-1 (1st part of domain f)

ATGGGATCAATGTACGTCACGTTGGTCGTAATCGGATTCGTATGGAACCTATCTGAACTCACCGAGA
GAGATCTCGCCAGCCTGAAATCTGCAATGAGGTCTCTACAAGCTGACGATGGGGTGAACGGTTATC
AAGCCATTGCATCATTCCACGGTCTCCCGGCTTCTTGTTCATGATGATGAGGGACATGAG

Fig. 10d

INTRON 2F (SEQ ID NO:151)

GTAAATAAAACGTCCAGTCATCGGAAACCCGCCAGATATATGGGTTTTTTTCTATTTAAACAAA
AAAGCAGAGACAAAAGATTATTTAAAGTCACATTTAACTTGATATCAGATCAATAGTTTGGCTAG
TTAGTGCTCTATATCCCTCAAATCCTTCGAATCTTTAAGCCTCGTGATATTTTGACAAACAGAGAA
GACTTAGTAGCCCAGACTTTCCCTTATTTTTTCCTGAAAATCTTAATACGGATATTAAATGGATTG
ATTCTGCAACCTACAACCATAGCCCATATGTTATTATTTTCAG

DOMAIN 2F-2 (2nd part of domain f)

ATTGCCTGTTGTATCCACGGAATGCCAGTATTCCCACACTGGCACAGGCTTTACACCCTGCAAATG
GACATGGCTCTGTTATCTCACGGATCTGCTGTTGCTATTCCATACTGGGACTGGACCAAACCTATC
AGCAAACCTGCCTGATCTCTTCACCAGCCCTGAATATTACGATCCTTGGAGGGATGCAGTTGTCAAT
AATCCATTTGCTAAAGGCTACATTAAATCCGAGGACGCTTACACGGTTAGGGATCCTCAGGACATT
TTGTACCACTTGCAGGACGAAACGGGAACATCTGTTTTGTTAGATCAAACCTCTTTTAGCCTTAGAG
CAGACAGATTTCTGTGATTTTGAGGTTCAATTTGAGGTCGTCCATAATGCTATTCACTACTTGGTG
GGTGGTCGACAAGTTTATGCTCTTTCTTCTCAACACTATGCTTCATATGACCCAGCCTTCTTTATT
CATCACTCCTTTGTTGACAAAATATGGGCAGTCTGGCAAGCTCTGCAAAAGAAGAGAAAGCGTCCC
TATCATAAAGCGGATTGTGCTCTTAACATGATGACCAAACCAATGCGACCATTGTCACACGATTTTC
AATCACAATGGATTTCACAAAATGCACGCAGTCCCCAACACTCTATTTGACTTTCAGGACCTTTTC
TACACGTATGACAACTTAGAAATTGCTGGCATGAATGTTAATCAGTTGGAAGCGGAAATCAACCGG
CGAAAAGCCAAACAAGAGTCTTTGCCGGGTTCTTCTACATGGCATTGGAAGATCAGCTGATGTA
CGATTTTGGATTGTGAAGACAGCTGACGACTGCCACGCATCTGGCATGATCTTTATCTTAGGAGGT
TCTAAAGAGATGCACTGGGCCTATGACAGGAACCTTTAAATACGACATCACCCAAGCTTTGAAGGCT
CAGTCCATACACCCTGAAGATGTGTTTGACACTGATGCTCCTTTCTTCATTAAAGTGGAGGTCCAT
GGTGTAACAAGACTGCTCTCCCATCTTCAGCTATCCCAGCACCTACTATAATCTACTCAGCTGGT
GAAG

INTRON 2F-2/2G (SEQ ID NO:152)

GTGAGAGAAACTATAATAGTGTATGTCGGCAAAAATGTGCTCATATCATGACTCTGTTGGCCGGT
GGTTGCTCTCCTCTCCTCCTCCACCACCACCGGTACCTCCACCTGTCAGGGCATCAATGTACCATG
AAAATGTCTACAATACTAGGCCTCCTGTAGAAGCACGTAAGATTTACATGGCCGGTTTGTAACTAG
TTTAAAGTGCTTCACAGTAACCAAACCAGTCTCTAAAGATTAATGTCTGTTTAAAATTTAATGCC
ACATTTTCAACTGACATATTCTTGCAATTAAGTACAAATGAAGTAGTATAAATTATCCACAAATAG
CGTGATGCACCACAAATATAAACCGAGTGCTTTTTTGGCATTCCCCACTTGTTCTGGCATGATCAC
ATCATAGATCTCGTTCATGAAGATACTGTTGGATGCTTTTTCCCAATATGCCCAATCTGTTAAAT
TATTTACACGACCGCAGTGTGTACTTTCATCACTCAGATCTTTACAATGTGTTTGTAAACGTTTACA
ATTAGCGTTATGATTGAAATATTACCCCTGCTACGTTAAATCACATTCACTCACTCATCTGATGT
ACTTTACAGGTCATACCGATGATCACGGCTCAG

DOMAIN 2G-1 (1st part of domain g)

ATCATATTGCTGGCAGTGGAGTCAGGAAAGACGTGACGTCTCTTACCGCATCTGAGATAGAGAACC
TGAGGCATGCTCTGCAAAGCGTGATGGATGATGATGGACCAATGGATTCCAGGCAATTGCTGCTT
ATCACGGAAGTCCTCCCATGTGTACATGCCTGATGGTAGAGACGTTGCATGTTGTACTCATG

INTRON 2G-1/2G-2 (SEQ ID NO:153)

GTCAGTATTCTCCAATATGTTTGACTAGTGTCTTGCTCATGTATCAACTATTTTAGGCAACGTTTT
TGATTGTTATGGTATTTTCATGATATGATTTTATTGCTACCTCTATACCCAAACAAAATGTTTTA

Fig. 10e

TCAACAATTGTTTGAGTTTTAATGCAAGAAAATTATCAGGAGTAGCGTGCAAAAATGACTGGAAGG
CATGGTGTACTTCTGTGTGTACATACAAGTGGGTAATGCCTTATTGAACTCGTAATCACTCGTTTC
AG

DOMAIN 2G-2 (2nd part of domain g)

GAATGGCATCTTCCCTCACTGGCACAGACTGTTTGTGAAACAGATGGAGGATGCACTGGCTGCGC
ATGGAGCTCACATTGGCATACTACTGGGATTGGACAAGTGCGTTTAGTCATCTGCCTGCCCTAG
TGACTGACCACGAGCACAATCCCTTCCACCAC

INTRON 2G-2/2G-3 (SEQ ID NO:154)

GTCAGTATTCTCCAATATGTTTGACTAGTGTCTTGCTCATGTATCAACTATTTTAGGCAACGTTTT
TGATTGTTATGGTATTTTCATGATATGATTTTATTGCTACCTCTATACCCAAACAAAATGTTTTA
TCAACAATTGTTTGAGTTTTAATGCAAGAAAATTATCAGGAGTAGCGTGCAAAAATGACTGGAAGG
CATGGTGTACTTCTGTGTGTACATACAAGTGGGTAATGCCTTATTGAACTCGTAATCACTCGTTTC
AG

DOMAIN 2G-3 (3rd part of domain g)

GGACATATTGCTCATCGGAATGTGGATACATCTCGATCTCCGAGAGACATGCTGTTCAATGACCCC
GAACACGGGTCAGAATCATTCTTCTATAGACAGGTTCTCTTGGCTCTAGAACAGACAGACTTCTGC
CAATTTGAAGTTCAGTTTGAAATAACACACAATGCAATCCACTCTTGGACTGGAGGACATACTCCA
TATGGAATGTCATCACTGGAATATACAGCATATGATCCACTCTTTTATCTCCACCATTCCAACACT
GATCGTATCTGGGCCATCTGGCAGGCACTCCAGAAATACAGAGGTTTTCAATACAACGCAGCTCAT
TGCGATATCCAGGTTCTGAAACAACCTCTTAAACCATTAGCGAGTCCAGGAATCCAAACCCAGTC
ACCAGAGCCAATTCTAGGGCAGTCGATTCATTTGATTATGAGAGACTCAATTATCAATATGACACA
CTTACCTTCCACGGACATTCTATCTCAGAACTTGATGCCATGCTTCAAGAGAGAAAGAAGGAAGAG
AGAACATTTGCAGCCTTCTGTTGCACGGATTTGGCGCCAGTGCTGATGTTTCGTTTGATGTCTGC
ACACCTGATGGTCATTGTGCCTTTGCTGGAACCTTCGCGGTACTTGTTGGGGAGCTTGAGATGCCC
TGGTCCTTTGAAAGATTGTTCCGTTACGATATCACAAAGGTTCTCAAGCAGATGAATCTTCACTAT
GATTCTGAGTTCCACTTTGAGTTGAAGATTGTTGGCACAGATGGAACAGAACTGCCATCGGATCGT
ATCAAGAGCCCTACCATTGAACACCATGGAGGAG

INTRON 2G/2H (SEQ ID NO:155)

GTATGTTTTGAGATCCACATAATCTTCTACCCTGTCTCATTTCTAATGCTCTTCAATACACAATTT
ATATAGCCTTTGAGCTTCAGATGTATTACGGACAGGCATTACAGTATACATGTAATATGGTTTTCT
GCTATTTGCAAAAATTGTGTCCTATCTCTGTTTCAGATCATCATGGCGGTGACACCTAG

DOMAIN 2H (SEQ ID NO:159)

GTCACGATCACAGTGAACGTCACGATGGATTTTTTCAGGAAGGAAGTCGGTTCCTGTCCCTGGATG
AAGCCAATGACCTTAAAAATGCACTGTACAAGCTGCAGAATGATCAGGGTCCCAATGGATATGAAT
CAATAGCCGGTTACCATGGCTATCCATTCTCTGCCCTGAACATGGTGAAGACCAGTACGCATGCT
GTGTCCACGGAATGCCTGTATTTCCACATTGGCACAGACTTCATACAATCCAGTTTGAGAGAGCTC
TCAAAGAACATGGTTCTCATTTGGGTCTGCCATACTGGGACTGGAC

Fig. 11aPrimary structure of the KLH2 proteinDOMAIN B

GLPYWDWTMPMSHLPELATSETYLDPVTGETKNNPFHHAQVAFENGVTSRNPDAKLFMKPTYGDHT
YLFDSMIYAFEQEDFCDFEVQYELTHNAIHAWVGGSEKYSMSSLHYTAFDPIFYLHHSNVDRLWAI
WQALQIRRGKSYKAHCASSQEREPLKPFASFSSPLNNEKTYHNSVPTNVYDYVGVVGLHYRYDDLQFG
GMTMSELEEYIHKQTQHDRTFAGFFLSYIGTSASVDIFINREGHDKYKVGSVVVLGGSKEMKWGFD
RMYKYEITEALKTLNVAVDGFSITVEITDVDGSPPSADLI PPPAII FDVVR

DOMAIN C

ADAKDFGHSRKIRKAVDSLTVEEQTSLRRAMADLQDDKTSGGFQQIAAFHGEPPKWCPSPAEKKFA
CCVHGMVFPWHRLTLVQGENALRKHGFTGGLPYWDWTRPMSALPHFVADPTYNDSVSSLEEDNP
WYHGHIDSVGHDTTRAVRDDLYQSPGFGHYTDIAKQVLLAFEQDDFCDFEVQFEIAHNFIHALVGG
NEPYSMSSLRYTTYDPIFFLHRSNTDRLWAIWQALQKYRGKPYNTANCAIASMRKPLQPFGLDSVI
NPDETREHSVPFRVFDYKNNFDYYESLAFNGLSIAQLDRELQRRKSHDRVFAGFLLHEIGQSAL
VKFYVCKHNVS DCDHYAGEFYILGDEAEMPWRYDRVYKYEITQQLHDLDLHVGDNFFLKYEAFDLN
GGSLGGSIFSQPSVIFEPAA

DOMAIN D

GSHQADEYREAVTSASHIRKNIRDLSERGEIESIRSAFLQIQKEGIYENIAKFHGKPGLCHEHDGHPV
ACCVHGMPTFPWHRLYLVLQVENALLERGSAAVAVPYWDWTEKADSLPSLINDATYFNSRSQTFDPN
PFFRGHIAFENAVTSRDPQPELWDNKDFYENVMLALEQDNFCDFEIQLELIHNALHSRLGGRAKYS
LSSLDYTAFDPVFFLHHANVDRIWAIWQDLQRYRKPYNEADCAVNEMRKPLQPFNNPELNSDSMT
LKHNL PQDSFDYQNRFRYQYDNLQFNHFSIQKLDQTIQARKQHDRVFA GFILHNI GTSAVVDIYIC
VEQGGEQNCKTKAGSFTILGGETEMPFHFDRLYKFDITSALHKLGVPLDGHGF DIKVDVRAVNGSH
LDQHILNEPSLLFVPGERKNIIY

DOMAIN E

DGLSQHNLVRKEVSSLTLEKHFLRKALKNMQADDS PDGYQAIASFHALPPLCPSPSAHRHACCL
HGMATFPQWHRLYTVQFEDSLKRHGSIVGLPYWDWLKPQSALPDLVTQETYEH LFSHKTFPNPFLK
ANIEFEGEGVTTERD VDAEHLFAKGNLVYNNWFCNQALYALEQENYCDFEIQFEILHNGIHSWVGG
SKTHSIGHLHYASYDPLFYIHHSQTDRIWAIWQALQEHRLSGKEAHCALEQMKDPLKPF SFGSPY
NLNKRTQEF SKPEDTFDYHRFGYEYDSLEFVGMSVSSLHNYIKQQQEADRVFAGFLLKGFGQSASV
SFDICRPDQSCQEAGYFSVLGGSSEMPWQFDRLYKYDITKTLKDMKLRYDDTFTIKVHIKD IAGAE
LDSDLIPTPSVLLEEGK

DOMAIN F

HGINVRHVGRNRIRMELSELTERDLASLKSAMRSLQADDGVNGYQAIASFHGLPASCHDDEGHEIA
CCIHGMPVFPWHRLYTLQMDMALLSHGSAVAIPYWDWTKPISKLPDLFTSPEYYDPWRDAVVNNP
FAKG YIKSEDAYTVRDPQDILYHLQDETGT SVLLDQTL LALEQTFCD FEFV VHNAIHYLVGG
RQVYALSSQHYASYDPAFFIHHSFVDKIWAVWQALQKKRKPYPHKADCALNMMTKPMRPFAHDFNH
NGFTKMHAVPNTLFD FQDLFYTYDNLEIAGMNVNQLEAEINRRKSQTRVFAGFLLHGIGRSADVRF
WICKTADDCHASGMIFILGGSKEMHWAYDRNFKYDITQALKAQSIHPEDVFD TDAPFFIKVEVHGV
NKTALPSSAIPAPTIISAGE

Fig. 11bDOMAIN G

DHIAGSGVRKDVTSLTASEIENLRHALQSVMDDDGPNGFQAIAAYHGSPPMCHMPDGRDVACCTHG
MASFPHWHRLFVKQMEDALAAHGAHIGIPYWDWTSASFSLPALVTDHEHNPFHGHIAHRNVDTSR
SPRDMLFNDPEHGSESFFYRQVLLALEQTDFCQFEVQFEITHNAIHSWTGGHTPYGMSSLEYTAYD
PLFYLHHSNTDRIWAIWQALQKYRGFQYNAAHCDIQVLKQPLKPFSESRNPNPVTRANSRAVDSFD
YERLNYQYDTLTFHGHISISELDAMLQERKKEERTFAAFLLHGFGASADVSDVCTPDGHCAFAGTF
AVLGGELEMPWSFERLFRYDITKVLKQMNLYDSEFHFELKIVGTDGTELPDRIKSPTIEHHGG

DOMAIN H (SEQ ID NO:158)

GHDHSEKHDGFFRKEVGSLSLDEANDLKNALYKLQNDQGPNGYESIAGYHGYPPFLCPEHGEGDQYAC
CVHGMPVFPWHRLHTIQFERALKEHGSHLGLPYWDW